UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

Thiago Gibbin Marconi

"Mapa Funcional em cana-de-açúcar utilizando marcadores moleculares baseados em SSR e SNP"

Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Genética e Biologia Molecular, na área de Genética Vegetal e Melhoramento.

Orientadora: Profa. Dra. Anete Pereira de Souza

Co-Orientador: Prof(a). Dr. Antonio Augusto Franco Garcia

Campinas – São Paulo – Brasil

Setembro de 2011

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

M333m

Marconi, Thiago Gibbin
Mapa funcional em cana-de-açúcar utilizando marcadores molecuares baseados em SSR e SNP / Thiago Gibbin Marconi. – Campinas, SP: [s.n.], 2011.

Orientador: Anete Pereira de Souza.
Co-orientador: Antônio Augusto Franco Garcia.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Cana-de-açúcar. 2. Poliploide. 3. Marcadores SSR. 4. Marcadores SNP. 5. Mapeamento funcional. I. Souza, Anete Pereira de, 1962-. II. Garcia, Antônio Augusto Franco. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Título em inglês: Functional genetic map of sugarcane using molecular markers based on SSR and SNP.
Palavras-chave em inglês: Sugarcane; Polyploid; SSR markers; SNP markers; Functional mapping.
Área de concentração: Genética Vegetal e Melhoramento.
Titulação: Doutor em Genética e Biologia Molecular.
Banca examinadora: Anete Pereira de Souza, Maria Lúcia Carneiro Vieira, Eugênio Cesar Ulian, Sérgio Furtado dos Reis, Roland Vencovsky.
Data da defesa: 13/09/2011.
Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 13 de setembro de 2011

BANCA EXAMINADORA

Profa. Dra. Anete Pereira de Souza (Orientadora)

Assinatura

Profa. Dra. Maria Lúcia Carneiro Vieira

Assinatura

Dr. Eugênio Cesar Ulian

Assinatura

Prof. Dr. Sérgio Furtado dos Reis

Assinatura

Prof. Dr. Roland Vencovsky

Assinatura

Prof. Dr. Marco Antonio Machado

Assinatura

Prof. Dr. José Baldin Pinheiro

Assinatura

Prof. Dr. Claudio Lopes de Souza Júnior

Assinatura

Dedicatória

Dedico aos meus queridos pais,

José Luiz e Maria Idê,

por todas as oportunidades oferecidas.

A minha querida esposa,

Cristina,

pela dedicação apoio durante este trabalho e sempre.

Agradecimentos

À Universidade Estadual de Campinas (UNICAMP), pelo apoio técnico e científico para o desenvolvimento de minha carreira acadêmica.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela bolsa de mestrado (2006/59361-9) concedida, Projeto Auxilio Pite: 2002/01167-1 e ao Programa Bioen, Projeto Temático: 2008/52197-4 por todos os recursos que tornaram possível a realização deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de doutorado (142128/2009-0) concedida e pela disponibilização de recursos para viabilizar este trabalho.

Ao Instituto Agronômico de Campinas (IAC) em especial ao Dr. Marcos Landell pela manutenção da parte experiemental de campo.

Especialmente a Profa. Dra. Anete Pereira de Souza, pela orientação, pela dedicação e apoio aos projetos, pelo carinho e conselhos que foram uma importante contribuição para a minha formação profissional e para a vida no qual tive o prazer de compartilhar desde a minha iniciação científica.

Ao Prof. Dr. Antonio Augusto Franco Garcia (ESALQ-USP) pela co-orientação, com o qual tive o prazer de trabalhar e compartilhar conhecimentos para a análise dos dados.

Aos colegas do laboratório de Análises Genética e Molecular/CBMEG e Barracão da Genética, agradeço tanto pelos momentos difíceis quanto aos momentos de descontração pois todo esse conhecimento adquirido com o grupo é que levarei comigo o resto da vida.

Aos amigos de trabalho do projeto cana, Karine, Luciana, Laura, Roberta, Tallis, Estela, Thiago, Melina e Benicio, pela fundamental ajuda no desenvolvimente do projeto, pela troca de conhecimentos e experiência, pela união e compreesão mesmo na horas de maior dificuldade e acima de tudo pela amizade.

Aos amigos do Laboratório de Genética de Estatística (ESALQ-USP), Gabriel, Marcelo, Maria Marta, Rodrigo, Renato, com o qual tive o prazer de conviver durante a finalização da tese, pelo grande auxílio nas análises e discussão dos dados e principalemente pela paciência.

Ao Juverlande, pela grande amizade e conselhos desde o meu início no laboratório.

Ao Carlão, Juverande e Pathy pelo apoio técnico prestado.

Às secretárias Sandra e Tânia do Centro de Biologia Molecular e Engenharia Genética (CBMEG) da Unicamp pela grande ajuda na parte financeira do projeto, pelo profissionalismo que me auxiliaram em todos os momentos que precisei.

Aos meus queridos irmãos, Rodrigo e Leandro, pelo apoio e carinho.

A família de minha esposa, em especial ao José Luiz e Lucia Helena, pelos conselhos no café da manhã e ensinamentos tanto da parte profissional quanto pessoal.

A minha esposa Cristina, pelo comprometimento e apoio principalmente durante a viagem a Austrália no qual juntos e ao mesmo tempo sozinhos do outro lado do mundo tivemos a oportunidade de um grande progresso na vida pessoal e profissional.

Especialmente aos meus pais, José Luiz e Maria Idê, no qual dedico este trabalho, pois sem todas as oportunidades oferecidas e sem todos os conhecimentos e valores pessoais por eles passados com certeza não seria possível mais esta conquista.

SUMÁRIO

Marconi, Thiago Gibbin. Mapa Funcional em cana-de-açúcar utilizando marcadores moleculares baseados em SSR e SNP. 2011. s.n.. Tese (Doutorado em Genética e Biologia Molecular) – Pós-graduação – Universidade Estadual de Campinas – UNICAMP – Campinas – SP - Brasil

RESUMO

A utilização dos marcadores moleculares em estudos de mapeamento genético e de QTLs (*Quantitative Trait Loci*) tem proporcionado um importante progresso no conhecimento da genética e da estrutura genômica da cana-de-açúcar. O projeto de sequenciamento de ESTs (*Expressed Sequence Tags*) do programa Genoma da FAPESP (SUCEST) identificou aproximadamente 43 mil *clusters* que representam os genes de cana-de-açúcar. Sabe-se que os ESTs apresentam grande potencial para serem utilizados no desenvolvimento de marcadores genético-moleculares. Tendo em vista os avanços possíveis no melhoramento genético da cana-de-açúcar com a construção de um mapa genético funcional a partir de ESTs de interesse, este trabalho teve como objetivos o mapeamento genético em uma população F_1 de cana-de-açúcar utilizando marcadores moleculares do tipo EST-SSRs (*Expressed Sequence Tags - Simple Sequence Repeats)* e SNP (*Single Nucleotide Polymorphism*), desenvolvidos a partir de seqüências ESTs homólogas a genes de interesse. Os SNPs desenvolvidos e mapeados demonstraram novos tipos de segregações possíveis de serem incorporadas ao mapeamento genético em cana-de-açúcar, representando avanços para a análise genética de poliplóides e possibilitando a saturação do mapa genético com marcadores completamente informativos. Os marcadores moleculares EST-SSRs e SNPs desenvolvidos e integrados ao mapa genético da cana-de-açúcar aumentaram sua resolução e também as possibilidades de mapeamento dos QTLs com maior precisão.

Palavras-chave: Cana-de-açúcar; Poliplóide; Marcadores SSR; Marcadores SNP; Mapeamento Funcional

Marconi, Thiago Gibbin. Functional genetic map of sugarcane using molecular markers based on SSR and SNP. 2011. s.n.. PhD Thesis (PhD in Genetics and Molecular Biology) – University of Campinas – UNICAMP – Campinas – SP - Brazil

ABSTRACT

The use of molecular markers in genetic mapping studies and QTL (Quantitative Trait Loci) has provided an important advance in knowledge of genetics and genomic structure of sugarcane. The sequencing project of ESTs (Expressed Sequence Tags) form FAPESP's Genome Program (SUCEST) identified approximately 43 000 clusters representing the sugarcane genes. It is known that the ESTs have great potential for use in the development of genetic molecular markers. Given the possible advances in genetic breeding of sugarcane with the construction of a functional genetic map from ESTs of interest, the aim of this study was the construction of a genetic map in a F_1 population of sugarcane using molecular markers EST-SSR (*Expressed Sequence Tags - Simple Sequence Repeats)* and SNP (Single Nucleotide Polymorphism) derived from ESTs sequences homologous to genes of interest. The developed and mapped SNPs demonstrated new types of segregation ratio that could be incorporated in the genetic mapping of sugarcane, representing advances for the genetic analysis of polyploid and allowing the saturation of the genetic map with fully informative markers. The EST-SSR markers and SNPs developed and integrated into the genetic map of sugarcane increased the resolution, coverage of the genome and also the possibilities of mapping QTLs with greater precision.

Keywords: Sugarcane; Polyploid; SSR markers; SNP markers; Functional mapping

PREFÁCIO

Os resultados obtidos durante o desenvolvimento deste trabalho de tese estão apresentados na forma de 5 artigos científicos. O primeiro artigo, intitulado *"Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs"*, publicado na revista *Plant Breeding* (125: 378-384; 2006), apresenta a caracterização de um conjunto de marcadores EST-SSRs que foram previamente desenvolvidos pelo grupo, bem como sua comparação a SSRs genômicos com relação ao poder discriminatório e conteúdo de polimorfismo. No segundo artigo, intitulado "*Functional integrated genetic linkage map based on EST-markers for a sugarcane (Saccharum spp.) commercial cross"*, publicado na revista *Molecular Breeding* (20:189-208, 2007), foram apresentados resultados da construção de um mapa genético integrado com marcadores EST-SSRs funcionais, utilizando a população F_1 de mapeamento obtida a partir do cruzamento SP80-180 x SP80-4966 (Centro de Tecnologia Canavieira, Piracicaba/SP). Nos terceiro e quarto artigos, sendo eles respectivamente "*Characterization of new polymorphic functional markers for sugarcane"*, publicado na revista *Genome* (52:191-209, 2009) e, "*Functional markers for gene mapping and genetic diversity studies in sugarcane"*, publicado revista *BMC Research Notes* (4:264, 2011) foram desenvolvidos centenas de marcadores EST-SSRs derivados do banco de dados SUCEST, para os quais foram realizadas análises do polimorfismo e do poder discriminatório em genotipagens envolvendo diferentes clones de cana-de-açúcar, bem como análises bioinformáticas, nas quais as sequências dos ESTs utilizados para o desenvolvimento das EST-SSRs foram categorizadas quanto aos seus processos biológicos e função molecular. Os resultados apresentados nesses dois últimos artigos constituíram um conjunto importante de marcadores moleculares funcionais,

os quais foram utilizados no mapeamento genético-molecular de cana-de-açúcar realizado nesse trabalho. O quinto e último artigo intitulado "*SNPs como uma nova ferramenta no mapeamento genético em espécies altamente poliplóides e aneuplóides: cana-de-açúcar"*, ainda não foi submetido à publicação. Nele relata-se o desenvolvimento de marcadores do tipo SNPs (*Single Nucleotide Polymorphism*) e sua integração no mapa genético previamente desenvolvido (com os marcadores EST-SSRs) utilizando a população F_1 de mapeamento obtida a partir do cruzamento IACSP 95-3018 x IACSP 93-3046 (Instituto Agronômico de Campinas - Centro de Cana, Ribeirão Preto/SP).

Durante o doutorado o autor dessa tese realizou estágio no *Centre for Plant Conservation Genetics,* (*Lismore/NSW/Australia*), sob orientação do Dr. Robert Henry e do Dr. Peter Bundock. No período de 8 meses de estágio, foram desenvolvidos 943 marcadores SNPs a partir de sequências expressas do banco de dados SUCEST. Os SNPs desenvolvidos foram genotipados na população F_1 (IACSP 95-3018 x IACSP 93-3046) utilizando um o equipamento *Sequenom*® para análise do DNA por espectrometria de massas.

INTRODUÇÃO

A cana-de-açúcar é uma espécie de grande importância para a agricultura brasileira e mundial, sendo cultivada principalmente em regiões de clima tropical e subtropical. Os principais produtos derivados são açúcar e etanol. No Brasil, cerca de 642 milhões de toneladas de cana devem ser processadas durante a safra 2011/2012 para produção de açúcar e álcool, sendo este valor 2,9% maior que o volume processado na safra 2009/2010 (Conab, 2011). O etanol, que vem sendo utilizado como biocombustível despertou grande interesse econômico-ambiental, pois trata-se de uma fonte de energia renovável competitiva em relação aos derivados do petróleo (Goldemberg, 2007; FAO, 2009).

Segundo Daniels e Roach (1987), a cana-de-açúcar é uma espécie alógama, da família Poaceae (Gramineae) e do gênero *Saccharum*, no qual ocorrem principalmente seis espécies: *S. officinarum* L. (2n = 80), *S. robustum* Brandes e Jeswiet ex Grassl (2n = 60-205), *S. barberi* Jeswiet (2n = 81-124), *S. sinense* Roxb. (2n = 111-120), *S. spontaneum* L. (2n = 40-128) e *S. edule* Hassk. (2n = 60-80). A espécie *S. officinarum* L. destaca-se por suas boas características agronômicas e altos teores de sacarose, sendo denominada como 'cana nobre' (Matsuoka et al., 1999b; Landell e Bressiani, 2008). Contudo, as variedades modernas de cana-de-acúcar correspondem a híbridos interespecíficos de elevada complexidade genética pois apresentam elevados níveis de ploidia e também aneuploidia (Heinz e Tew, 1987; Matsuoka et al., 1999b; Landell e Bressiani, 2008). Acredita-se que dentre os genótipos cultivados haja contribuição, em proporções variáveis, do genoma de diferentes espécies do gênero *Saccharum*, principalmente, *S. officinarum* e *S. spontaneum*. Neste contexto, considera-se que a cana-de-açúcar é a espécie cultivada de maior complexidade genética (Manners et al., 2004).

A grande importância da cana-de-açúcar e seus derivados para os países tropicais e subtropicais tem resultado, recentemente, em investimentos na pesquisa e na obtenção de tecnologias para auxiliar nos programas de melhoramento e no entendimento da genética dessa cultura. Esse fato se deve, principalmente, à complexidade do genoma da cana e ao longo tempo necessário (15 anos) para a seleção de novos cultivares mais produtivos através de métodos de melhoramento tradicionais (D'Hont et al., 2008). Entre as ferramentas derivadas da genômica que vem sendo utilizadas no melhoramento genético estão os marcadores moleculares (revisto por Grivet e Arruda, 2002).

Os marcadores moleculares têm se mostrado particularmente úteis para auxiliar programas de melhoramento em diversas culturas (Francia et al., 2005), em cana-de-açúcar para estudos de diversidade genética (Lima et al., 2002), obtenção de mapas genéticos (Aitken et al., 2005; Aitken et al., 2007; Raboin et al., 2006; Garcia et al., 2006; Oliveira et al., 2007) e mapeamento de QTL (Quantitative Trait Loci) (MCintyre et al., 2006; Al-Janabi et al., 2006; Piperidis et al., 2008; Aitken et al., 2008). Os trabalhos desenvolvidos com marcadores moleculares em cana-de-açúcar, especialmente aqueles referentes à construção de mapas genéticos de ligação e mapeamento associativo, são importantes para aumentar o entendimento da arquitetura do genoma da cana, auxiliando na diminuição do tempo de obtenção de cultivares melhorados, através da identificação de regiões no genoma associados a caracteres de interesse econômico.

O projeto Genoma SUCEST (Sugarcane Expressed Sequence Tag) da FAPESP produziu um extenso banco de dados de etiquetas de seqüências expressas (ESTs) para cana, com a identificação de 237.954 ESTs agrupados em cerca de 43.000 clusters (Vettore et al., 2001, 2003). A maioria dos clusters representa genes envolvidos em importantes processos

metabólicos, tais como fotossíntese, metabolismo de carboidratos, transporte de açúcares, metabolismo de aminoácidos, mecanismo de resposta aos estresses bióticos e abióticos, entre outros. Esse banco de dados serviu como fonte para o desenvolvimento de marcadores moleculares EST-SSRs (Simple Sequence Repeats) (Pinto et al., 2004; Oliveira et al., 2009) que foram utilizados juntamente com marcadores SSRs genômicos e RFLP (Restriction Fragment Length Polymorphism) na construção do primeiro mapa genético funcional para cana-de-açúcar (Oliveira et al., 2007). Apesar da abundância de microssatélites no banco de dados do SUCEST, eles foram encontrados em apenas 5% do total de sequências (Pinto et al., 2004), de modo que não foi possível desenvolver esse tipo de marcador para 95% dos ESTs do SUCEST.

A geração de um mapa genético para cana-de-açúcar requer um esforço muito maior quando comparado àquele dispendido com o mesmo objetivo para outras espécies, devido ao genoma da cana ser altamente poliplóide. Devido a esse fato, os mapas para cana-de-açúcar gerados até o momento são pouco densos, sendo que mais marcadores são necessários para a obtenção de mapas saturados e que possibilitem a identificação de maior número de QTLs.

Os SNPs são responsáveis pela maior parte da variação presente no genoma das espécies e explicam a ocorrência de muitas características de interesse em plantas (Umemoto et al., 2004; Varshney et al., 2007). Sendo assim, nesse trabalho objetivou-se o desenvolvimento de marcadores moleculares funcionais do tipo SNPs para cana-de-açúcar, os quais estão localizados em ESTs que apresentam similaridade com genes de interesse. Estes marcadores estão sendo integrados ao mapa genético-molecular previamente desenvolvido para a população F_1 de cana-de-açúcar, obtida do cruzamento IACSP 95-3018 x IACSP 93-3046. Para a integração dos SNPs, foi necessário o desenvolvimento de novas ferramentas para a análise desses marcadores, visto que nunca antes haviam sido analisados para a integração em mapas

genéticos de organismos poliplóides alógamos, como é a cana-de-açúcar. Esta espécie é um poliplóide interespecífico bastante complexo quando comparado a outros. Cada loco de uma planta de cana-de-açúcar pode ter número variável de cópias no genoma, seja por duplicação, seja pela presença de diferentes cromossomos homeólogos de origem multiespecífica. Além disso, o mesmo loco pode ter número de cópias variável no genoma de diferentes indivíduos.

Os resultados desse trabalho serão de grande importância para possibilitar a utilização de marcadores moleculares no estudo genético de organismos poliploides, em geral. Novos tipos de segregação observados e analisados poderão ser integrados ao mapa genético de poliplóides, melhorando a saturação do mapa genético e aprimorando a descoberta de QTLs. Dessa forma, será possível melhorar o entendimento da arquitetura genética da cana-de-açúcar e acelerar o melhoramento genético dessa cultura, bem como de outros organismos de origem poliplóide.

REVISÃO BIBLIOGRÁFICA

1 Aspectos Gerais da cultura da cana-de-açúcar

1.1 Importancia econômica

A principal fonte de açúcar provém da cana-de-açúcar, responsável por 70% da produção mundial, e da beterraba açucareira, com uma contribuição de 30% (Ming et al., 2006). O Brasil é o maior produtor e exportador de açúcar do mundo, sendo responsável, em termos mundiais, por aproximadamente 20% da produção e 40% das exportações. A produção nacional em 2007/08 foi de 31 milhões de toneladas. Cerca de dois terços do açúcar produzido no Brasil (18,6 milhões de toneladas) destinaram-se à exportação (UNICA, 2008).

Outro importante derivado da cana-de-açúcar é o etanol, do qual 30% da produção mundial origina-se da cana-de-açúcar a partir da fermentação do açúcar (UNICA, 2008). O etanol está se tornando uma das principais fontes alternativas de energia, principalmente devido ao aquecimento global e aos preços instáveis e elevados do petróleo (Ortolan, 2006). Os Estados Unidos e o Brasil são responsáveis pela maior parte da produção de etanol do mundo, 42% e 37%, respectivamente (UNICA, 2008). Apesar dos Estados Unidos possuírem atualmente a maior produção de etanol do mundo, ela é principalmente baseada no milho como matéria-prima. O milho, em comparação com a cana-de-açúcar, possui um menor rendimento na produção de etanol, além de afetar diretamente o preço e a disponibilidade de alimentos de ração animal (English et al., 2005). A exportação brasileira de etanol durante a safra 2008/2009 teve um aumento 30% em volume em relação à safra anterior 2007/2008, atingindo um total de 4,7 bilhões de litros. Ao compararmos os números atuais referentes à exportação brasileira de etanol em relação à safra de 2004/2005 observa-se um incremento de 90%, o que evidencia o

crescimento exponencial da produção de etanol no Brasil.

A crescente demanda pelo etanol no mundo é decorrente do incentivo ao uso de matrizes energéticas renováveis, diminuindo a dependência do petróleo proveniente de áreas de instabilidade política e da volatilidade de seus preços. No Brasil, com o advento do carro *flex fuel,* o consumo interno aumentou significativamente nos últimos anos. Na safra de 2008/2009 foram produzidos 18 bilhões de litros, representando um incremento de 27% em relação à safra anterior. A venda de carros *flex fuel* corresponde a 92% do total de carros vendidos atualmente no Brasil, o que resulta em uma previsão muito otimista para o aumento de consumo de etanol no mercado interno (UNICA, 2011).

Nos últimos anos, grupos de pesquisa (públicos e privados) ao redor do mundo estão empenhados no estabelecimento de tecnologias para a conversão da biomassa proveniente do processamento da cana-de-açúcar em açúcares fermentáveis, resultando no etanol de segunda geração, para o aumento na produção de etanol (Demirbas, 2001). Embora essa tecnologia ainda não possa ser aplicada em escala industrial, ela tem grande importância nesse contexto de crescente demanda (Badger, 2002). Por todos esses motivos o álcool produzido da cana-de-açúcar é o principal propulsor do setor sucroalcooleiro e, atualmente, movimenta grandes montantes de investimentos (Boyle et al., 2008).

Em vista do cenário internacional, a demanda por açúcar e etanol tende a crescer e o Brasil possui uma posição privilegiada para atender a estas necessidades. O país tem duas regiões produtoras, Centro-Sul e no Norte-Nordeste, com safras alternadas, podendo manter sua presença no mercado mundial ao longo de todo o ano. O Brasil possui uma avançada tecnologia de produção de álcool, além de ter o menor custo de produção do mundo e ainda possuir potencial de expansão de área plantada e de produtividade.

1.2 Origem e características das espécies do gênero *Saccharum*

Acredita-se que o centro de diversidade das espécies do gênero *Saccharum*, segundo reportam alguns autores, está localizado na região das ilhas do Pacífico Sul, incluindo a Índia, Indonésia, Nova Guiné, China, Ilha Polinésia e Ilhas Fiji (Brandes, 1956; Mukherjee, 1957; Daniels e Roach, 1987). Em meados do século XV, durante suas expedições, os portugueses e espanhóis propagaram esta cultura nas suas colônias da região do Atlântico. Durante os séculos XVI e XVII a cultura de cana-de-açúcar estava intimamente relacionada à colonização européia, principalmente no Brasil e no Caribe.

Martim Afonso de Souza foi quem oficialmente trouxe a primeira muda de cana-de-açúcar ao Brasil em 1532, iniciando o seu cultivo na capitania de São Vicente. A partir do desenvolvimento da cultura em Pernambuco e Bahia, os engenhos de açúcar se multiplicaram pelo Brasil, onde se adaptou muito bem. Cinquenta anos após a chegada da cana-de-açúcar, o Brasil já era o maior produtor de açúcar do mundo (Machado et al., 1987)

A cana-de-açúcar caracteriza-se por ser uma planta herbácea, alógama e cultivada principalmente em regiões tropicais e subtropicais. Pertence a família das gramíneas (Poaceae), tribo Andropogoneae e gênero *Saccharum* (Tabela 1).

Tabela 1. Classificação taxonômica da cana-de-açúcar (Daniels e Roach, 1987)

FAMÍLIA	SUBFAMÍLIA	TRIBO	GRUPOS DE SUBTRIBO	SUBTRIBO	GÊNERO
Gramineae	Pooideae				
	Panicoideae	Paniceae			
		Andropogoneae	Dimeriinae Hack		
			Saccharinae Benth	"Saccharastrae"	Imperata Cyr
					Eriochrysis P. Beauv.
					Ecooilopus Steud
					Spodiopogon Trin.
					Miscanthidium Stapf.
					Erianthus Michx sect. Ripidium Henrard
					Miscanthus Anderss. sect. Diandra Keng
					Sclerostachya (Hack.) A. Camus
					Narenga Bor.
					Saccharum L.
			Germainiinae WD Clayton		
			Arthraxoninae Benth.		
			Andropogoninae Presl.		
			Anthistiriinae Presl.		
			Ischaeminae Presl.		
			Rottboeliinae Presl.		
			Tripsacinae Presl.		
			Coicinae		
			Chionachninae WD Clayton	"Eulaliastrae"	
	Chloridoideae				
	Arundinoideae				
	Bambusoideae				

"Complexo *Saccharum*"

No complexo *Saccharum* destacam-se cinco principais espécies: *S. spontaneum*, *S. robustum*, *S. officinarum*, *S. barberi*, *S. sinense* e *S. edule* descritas a seguir.

S. spontaneum é considerada uma espécie selvagem com número cromossômico elevado. Possui alta adaptabilidade a regiões com características edafoclimáticas extremamente adversas, relacionadas à temperatura, umidade e tipo de solo. Por este motivo, *S. spontaneum* apresenta vasta distribuição geográfica, que vai do Japão ao leste da África, passando pelo sudeste da Ásia, Índia, Oriente Médio e a bacia mediterrânea (Brandes et al., 1938). Por ser uma espécie selvagem, apresenta um baixo teor de sacarose. Em contrapartida, possui resistência a pragas e doenças, capacidade de rebrota de soqueira, vigor e grande adaptabilidade (Naidu e Sreenivasan, 1987). O número de cromossomos varia de 2n = 40 a 128, sendo os citótipos mais frequentes os múltiplos de 8, sugerindo que o número básico desta espécie é x = 8 (Panje e Babu, 1960; Naidu e Sreenivasan, 1987; Burner, 1991). De fato, D'Hont et al (1996) confirmaram este número por hibridização *in situ* de rDNAs, relatando também que o nível de ploidia varia entre 8 e 12, segundo os citótipos estudados.

S. robustum também é considerada uma espécie selvagem no gênero *Saccharum*. As principais diferenças em relação a *S. spontaneum* referem-se a ausência de rizomas, presença de inflorescência grande, haste mais espessa e maior altura (Stevenson, 1965). Foram encontrados dois citótipos euplóides, com 2n = 60 ou 2n = 80, e citótipos aneuplóides, variando de 2n = 63 a 205. Estes clones aneuplóides podem corresponder a híbridos naturais entre clones de *S. robustum* ou entre *S. robustum* e outras espécies (Price, 1957; 1965).

S. officinarum é a espécie mais parecida com as variedades comerciais encontradas hoje em dia, ela é conhecida como "cana nobre" por apresentar qualidades agronômicas e industriais correspondentes aos mais importantes critérios de seleção, tais como: colmos

grossos, alto teor de sacarose e baixo conteúdo de fibras e amido (Bremer, 1961). Acredita-se que esta espécie originou-se a partir de formas 2n = 80 da espécie selvagem *S. robustum* na Nova Guiné, sendo este seu centro de origem (Brandes et al., 1938). A partir da Nova Guiné, os clones de *S. officinarum* se dispersaram para as ilhas vizinhas e, em seguida, para várias regiões que se tornaram os atuais países produtores de açúcar. *S. officinarum* é uma espécie euplóide, com 2n = 80 cromossomos (Bremer, 1930; Li e Prince, 1967; Price e Daniels, 1968). Acredita-se que alguns poucos clones de *S. officinarum* que não apresentam este número cromossômico se originaram de hibridação com outras espécies (Bremer, 1924). *S. officinarum* é uma espécie octaplóide, apresentando um número cromossômico básico de x = 10 (D'Hont et al., 1996).

S. barberi* e *S. sinense são espécies que já foram cultivadas respectivamente no norte da Índia e no sul da China (Stevenson, 1965). Acredita-se que sua origem deve-se a hibridações naturais entre as espécies *S. officinarum* e *S. spontaneum*, nas suas respectivas regiões de origem (Price e Daniels, 1968). Ambas as espécies seriam híbridos de primeira geração, já que poucas recombinações interespecíficas entre os cromossomos foram encontradas em estudo realizados com hibridização *in situ* (D'Hont et al., 1996). As principais diferenças em relação a *S. officinarum* são referentes às suas características florais e ao elevado teor de fibras e rusticidade. Um grande número de clones destas espécies é resistente às principais doenças da cana-de-açúcar (Daniels e Roach, 1987). Seu número cromossômico varia entre 2n = 81 a 124 em *S. sinense* e 2n = 111 a 120 em *S. barberi*. Observações citológicas mostram meioses extremamente irregulares (Sreenivasan et al., 1987).

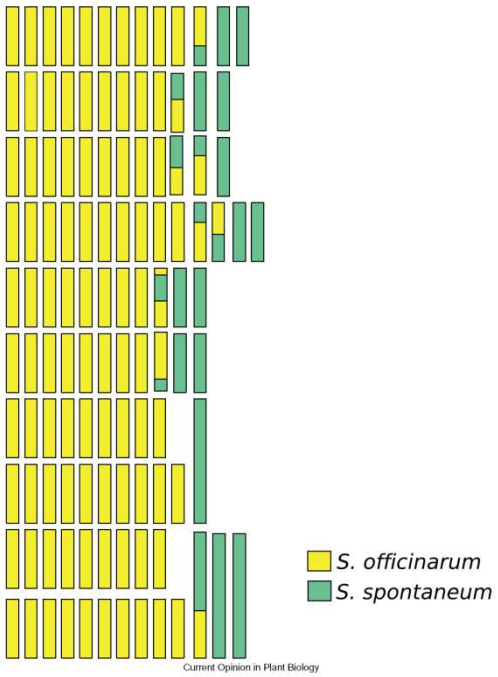
S. edule é um grupo com menor representatividade que ocorre principalmente na Nova Guiné e ilhas vizinhas. Sua principal característica é a esterilidade. A origem dessa espécie deu-se provavelmente a partir da espécie *S. robustum* por hibridação interespecífica, sendo *S.*

robustum como doadora do gameta feminino (Grivet et al., 2004). Os perfis moleculares mitocondriais e cloroplásticos dos clones de *S. edule* são semelhantes aos perfis destas organelas mais frequentemente observados na espécie *S. robustum* (D'Hont et al., 1993; Sobral et al., 1994). Os clones de *S. edule* constituem uma série de poliplóides com 2n = 60, 70 ou 80 cromossomos (Roach, 1972).

1.3 Cultivares modernos de cana-de-açúcar

Estimulados pelo surgimento de novas doenças, no início do século XX os programas de melhoramento da região da Índia e de Java produziram híbridos interespecíficos resistentes à doenças e mais adaptados às condições ambientais encontradas. Na Índia foram conduzidos cruzamentos interespecíficos entre *S. officinarum*, *S. spontaneum* e *S. barberi*. Simultaneamente, em Java foram realizados cruzamentos interespecíficos utilizando *S. officinarum* e *S. spontaneum*. A partir desses cruzamentos surgiu o termo "nobilização", que significa o cruzamento de uma espécie rica em açúcar (cana nobre) com uma espécie com características mais robustas (cana selvagem), e realização de diversos retrocruzamentos do híbrido com a variedade nobre para recuperação das características desejáveis. Desta forma, a maior parte do genoma de um híbrido de cana-de-açúcar vem de *S. officinarum* (80%) (Figura 1) (D'Hont et al., 1996; Grivet e Arruda., 2002), sendo portanto, o objetivo da nobilização a incorporação de vigor e resistência à doenças aos clones cultivados. Este foi um passo muito importante na produção das variedades modernas encontradas hoje em dia, as quais possuem a característica de alta produção de açúcar e resistência a pragas e doenças (Roach, 1972; Ming et al., 2006).

Figura 1. Representação esquemática do genoma das variedades comerciais atuais de cana-de-açúcar. Cada barra representa um cromossomo e cada cor representa uma origem, caracterizando a sua natureza poliplóide, aneuplóide e interespecífica (Grivet et al, 2002).



Os primeiros híbridos interespecíficos gerados a partir da nobilização foram utilizados maciçamente em cruzamentos em todos os programas de seleção através do mundo. São exemplos de variedades obtidas desses cruzamentos: POJ2878, POJ2725, POJ213 de Java e as variedades Co281, Co290 de Coimbatore, as quais constituem a base genética das atuais variedades comerciais de cana-de-açúcar no mundo todo.

O processo de nobilização gerou híbridos de origem poliplóide e aneuplóide, apresentando número cromossômico variável de 2n=100 a 130 (D'Hont et al., 1996). A aneuploidia é comum em espécie de origem poliplóide natural e híbrida (Sybenga, 1992), e casos de instabilidade cromossômica ocorrem predominantemente em plantas com origem poliplóide apresentando elevado número cromossômico como é o caso da cana-de-açúcar

(D'Hont et al., 1996), porém, a cana-de-açúcar apresenta a vantagem da propagação vegetativa para a manutenção da estabilidade cromossômica.

O pareamento dos cromossomos durante a meiose é complexo, porém a disomia é muito improvável indicando que todos os cromossomos homólogos sofrem recombinação (Grivet et al., 1996; Hoarau et al., 2001). Levando este fato em consideração pode-se inferir que um gene é representado em aproximadamente 10 cópias, cada uma podendo corresponder a um haplótipo diferente. Entre os 10 alelos, de 8 e 9 foram provavelmente herdados de *S. officinarum* e 1 ou 2 do *S. spontaneum* (Grivet et al., 2003).

2 Melhoramento genético de cana-de-açúcar

Apesar da complexidade genômica da cana-de-açúcar, a obtenção de novas variedades melhoradas é em parte facilitada pela existência de mecanismos de propagação vegetativa que fixam o genótipo desejado. Os programas de melhoramento genético são baseados no cruzamento de indivíduos com características agronômicas desejáveis que levam à obtenção de populações segregante F_1 com variabilidade genética. Estas populações sofrem diversas etapas de seleção e propagação vegetativa dos genótipos de interesse. Cada programa de melhoramento possui suas especificidades referentes aos métodos e objetivos finais almejados. Dessa forma, diferentes estratégias podem ser adotadas. Mesmo considerando dessas especificidades, a obtenção de uma variedade melhorada pode demorar de 10 a 12 anos (Matsuoka et al., 1999b; Landell e Bressiani, 2008).

Os principais fatores que influenciam no resultado dos programas de melhoramento são: escolha adequada dos genitores; instalação de experimentos com boa precisão experimental; escolha correta dos caracteres agronômicos de interesse e realização de avaliações periódicas.

Entre as características de interesse a serem selecionadas no programa de melhoramento de cana-de-açúcar estão: teor de sólidos solúveis (BRIX), teor de sacarose, diâmetro e número de colmos, perfilhamento, teor de fibras, resistência ao acamamento e florescimento, precocidade e resistência a pragas e doenças (Matsuoka et al., 1999a; Matsuoka et al., 1999b). Esses caracteres são geralmente controlados por muitos locos, sendo denominados de caracteres quantitativos, cujos locos se encontram localizados em QTLs (*Quantitative Trait Loci*).

De forma geral, as principais etapas de um programa de melhoramento para cana-de-açúcar são: a) cruzamento; b) seleções iniciais; c) seleção utilizando experimentos com repetições; d) ensaios finais de competição.

a) Cruzamento

A escolha dos genitores é um passo importante no programa de melhoramento. A partir da variabilidade gerada no cruzamento espera-se encontrar elevada probabilidade de aparecimento de genótipos superiores, alguns deles com grandes chances de se tornarem uma futura variedade comercial (Matsuoka et al., 1999a). Devido à estreita base genética do germoplasma utilizado para o melhoramento genético da cana-de-açúcar, durante o programa de melhoramento, há o risco de surgirem problemas de depressão por endogamia. Muitos programas de melhoramento usam também como critério de seleção dos genitores a genealogia e/ou divergência genética entre eles, obtida através de marcadores moleculares (Lima et al., 2002), para contornar o problema da base genética estreita do germoplasma.

Após definidos os genitores é necessário definir a estratégia para a obtenção da população segregante. Duas estratégias são muito utilizadas em cana-de-açúcar: cruzamentos biparentais e policruzamentos (Matsuoka et al., 1999a; Matsuoka et al., 1999b; Landell e

Bressiani, 2008). Nos cruzamentos biparentais a população segregante é obtida utilizando-se apenas dois genitores, um como genitor masculino e outro feminino, sendo este o cruzamento mais utilizado. Nos policruzamentos são utilizados mais de dois genitores, sendo apenas um como genitor feminino e dois ou mais como genitores masculinos simultaneamente. A principal vantagem do policruzamento é o grande número de sementes obtidas quando comparado aos cruzamentos biparentais, bem como a maior variabilidade dentro da população F_1 gerada.

b) Seleções iniciais

Com a geração de uma população segregante, o principal desafio é encontrar os genótipos superiores dentre os milhares de indivíduos gerados no cruzamento. Entretanto, a seleção de caracteres que apresentam alta influência ambiental pode ser prejudicada se os ensaios experimentais para avaliação do valor fenotípico de cada genótipo não forem realizados adequadamente. Desta forma, o processo de seleção ocorre inicialmente para caracteres com alta herdabilidade e baixa intensidade de seleção. Ao longo das etapas de seleção, a disponibilidade de colmos por genótipos aumenta gradativamente, o que permite a instalação de experimentos com níveis de precisão cada vez mais elevados, com intensidade de seleção crescente. O tamanho das parcelas e o número de repetições variam de acordo com a disponibilidade de colmos e características específicas sob seleção nos programas de melhoramento.

A cana-de-açúcar apresenta uma vantagem perante outras espécies vegetais como milho, sorgo, soja, feijão, pois possibilita a propagação vegetativa dos genótipos a cada ciclo de seleção. Tal fato torna possível explorar a herdabilidade de características de interesse em sentido amplo, ou seja, além das variações aditivas é possível verificar a existência de dominância e interações epistáticas (Hill et al., 2008). De acordo com os resultados

apresentados por Skinner et al. (1987) e Matsuoka et al. (1999b), as fases iniciais de seleção são mais efetivas para características como o teor de açúcar (brix) e resistência a carvão e ferrugem. Contudo, os resultados devem ser interpretados com cautela, pois estimativas de herdabilidades podem variar de acordo com a população e as condições experimentais utilizadas (Hill et al., 2008).

Na condução do plantio para seleção nos estágios iniciais é possível utilizar algumas estratégias como plantio em *bunch*, individual ou em famílias para ajudar o melhorista, visto que não é possível realizar o plantio utilizando delineamentos estatísticos.

O plantio em *bunch* (Mangelsdorf, 1953) constitui-se no plantio de 5 a 10 plântulas por cova, permitindo que a seleção natural atue eliminando genótipos inferiores. Uma vantagem dessa metodologia é a possibilidade de avaliar uma maior quantidade de colmos simultaneamente em uma mesma área. Segundo Skinner et al. (1987) há exemplos de sucesso obtido utilizando este tipo de plantio. Em contrapartida Matsuoka et al. (1999b) apontam como desvantagem dessa metodologia, a impossibilidade de se avaliar o perfilhamento, o qual é correlacionado positivamente com a capacidade de brotação e produtividade.

A seleção de plantas individuais consiste no plantio dos genótipos num espaçamento maior, evitando competição entre as plantas. Este tipo de estratégia é recomendado para a seleção de caracteres com alta herdabilidade, tais como Brix, resistência ao florescimento e as doenças (Matsuoka et al., 1999b).

A estratégia de seleção de famílias é utilizada em situações nas quais existe vários cruzamento a serem avaliados em conjunto, sendo eles biparentais ou policruzamentos. As progênies dos diferentes cruzamentos são avaliadas de acordo com o comportamento médio dos indivíduos. Nesta abordagem explora-se o conceito de herdabilidade em nível de médias, o

que permite a seleção de caracteres quantitativos em etapas iniciais do processo que não poderiam ser avaliados em seleção de plantas individuais.

Ao se considerar a população segregante como um todo, é possível realizar estudos de interação genótipos por ambientes (Hogarth e Bull, 1990; Bull et al., 1992; Jackson et al., 1995a, 1995b; Bressiani et al., 2002; Landell e Bressiani, 2008), além de identificar cruzamentos superiores (Skinner et al., 1987) possibilitando também facilidades na implementação dos experimentos, como a mecanização do processo de seleção (Landell e Bressiani, 2008). Em função disso, vários programas de melhoramento preferem utilizar esta estratégia, escolhendo as famílias que possuem médias fenotípicas superiores para as avaliações posteriores. No entanto, tal procedimento pode ocasionar o descarte de genótipos superiores se eles estiverem presentes em famílias com baixa média e variância elevada (Matsuoka et al., 1999b).

c) Seleção utilizando experimentos com repetições

Os genótipos selecionados são então comparados entre si com base em experimentos utilizando delineamentos estatísticos apropriados. Uma característica destes experimentos é o elevado número de plantas em avaliação e o reduzido número de colmos para a obtenção de repetições do experimento em outros ambientes. Um delineamento apropriado para esta etapa é o delineamento de blocos casualizados completos aumentados (Federer, 1956).

Com a utilização de delineamentos experimentais há um maior controle das variáveis ambientais, o que já possibilita o início da seleção para características influenciadas pelo ambiente, contempladas na etapa anterior, como por exemplo, avaliações para capacidade de rebrota. Ao final de cada experimento reduz-se o número de genótipos e aumenta-se a disponibilidade de colmos. Estes colmos são utilizados para o plantio de parcelas maiores, de maior número de repetições e para a implantação de experimentos em outros ambientes. Essas

modificações nas avaliações resultarão em maior precisão experimental possibilitando avaliações de caracteres com herdabilidades reduzidas.

d) Ensaios finais de competição

Após uma série de etapas de seleção torna-se possível a realização de experimentos com elevado nível de precisão experimental, o que permite avaliar caracteres de baixa herdabilidade, utilizando cultivares comerciais como testemunhas, visando identificar genótipos superiores em relação aos padrões comerciais. Nesta etapa são realizados testes de estabilidade e adaptabilidade, para selecionar genótipos mais aptos a determinadas condições ambientais.

Os programas de melhoramento de cana-de-açúcar no Brasil usualmente realizam experimentos utilizando o delineamento em blocos ao acaso (Matsuoka et al., 1999b; Bidoia, 2008) repetidos em diversos ambientes, com avaliações em pelo menos dois cortes. Os genótipos que se destacam nesta fase são recomendados como novas variedades. Nessa fase há grande envolvimento das usinas, sendo que os experimentos são instalados e avaliados em condições de cultivo muito próximas àquelas que são encontradas nas práticas agronômicas das usinas, possibilitando que as variedades posteriormente selecionadas sejam imediatamente utilizadas pela usina.

3 Marcadores moleculares em cana-de-açúcar

Os marcadores moleculares são ferramentas valiosas no estudo de genomas complexos como o da cana-de-açúcar (Daugrois et al., 1996). A utilização de marcadores na seleção de características agronômicas desejáveis durante os estágios iniciais do melhoramento, ou mesmo na escolha de genitores em um cruzamento, poderia reduzir significantemente o tempo de desenvolvimento de novas variedades comerciais.

O desenvolvimento de bancos de dados de marcadores moleculares polimórficos e robustos, ligados a características agronômicas de interesse, é imprescindível para seu uso associado aos programas de melhoramento. Com o auxílio dos marcadores moleculares foram efetuados avanços visando o maior conhecimento da complexidade genética e genômica das variedades modernas de cana-de-açúcar. As informações obtidas por meio de marcadores moleculares, sobre a diversidade entre as cultivares (LU et al., 1994a; Jannoo et al., 1999; Lima et al., 2002), as principais variedades utilizadas como genitores (Glaszmann et al., 1990; D'Hont et al., 1993; Lu et al., 1994b; Burnquist et al., 1995; Jannoo et al., 1999; Nair et al., 1999), e em outros gêneros do complexo *Saccharum* (AL-Janabi et al., 1994; Besse et al., 1996; Alix et al., 1998, 1999) têm grandemente auxiliado os melhoristas na escolha da melhor estratégia para a exploração da variabilidade genética existente.

Os marcadores moleculares também têm sido utilizados como uma ferramenta na identificação de variedades, no controle da progênie e no monitoramento de introgressão (D'Hont et al., 1995; Harvey et al., 1998; Cordeiro et al., 2000; Piperidis e D'Hont, 2001).

O emprego da seleção assistida por marcadores (SAM) nos programas de melhoramento de cana-de-açúcar ainda está distante de ser realidade, embora este não seja o caso para culturas como o milho, arroz, cevada e soja entre outras (Francia et al., 2005). A SAM fundamenta-se no conceito de que é possível inferir a presença de um gene a partir de um marcador intimamente ligado a ele. Quando o marcador se encontra muito longe da região de interesse, a possibilidade de ambos serem transmitidos aos indivíduos da progênie é reduzida devido aos eventos de recombinação. Sendo assim, a existência de uma forte ligação entre a característica de interesse e o marcador é pré-requisito deste tipo de seleção (Kumar, 1999; Francia et al., 2005).

3.1 Microssatélites

Microssatélites ou SSRs (*Simple Sequence Repeats*) são regiões do genoma que contém pequenas sequências de nucleotídeos apresentando de uma a seis bases de comprimento repetidas em *tandem.* O número total das repetições em tandem pode variar de um indivíduo a outro, o que resulta em polimorfismo, o qual é extremamente útil em estudos genéticos. O polimorfismo dos marcadores SSRs é revelado pela reação em cadeia da polimerase (PCR – Polymerase Chain Reaction) utilizando como molde o DNA genômico total e dois *primers* específicos que flanqueiam a região que contém o SSR.

Os SSRs são amplamente utilizados em estudos de mapeamento genético, *fingerprinting*, estudos evolutivos e de diversidade populacional, tanto em plantas como em animais (Bell e Ecker, 1994; Yang et al., 1994; Russell et al., 1997; Petren et al., 1999; Senior et al., 1996). Algumas características que contribuem para o amplo uso dos SSRs são atribuídas a sua natureza multialélica, herança codominante e transferibilidade entre espécies (Griffiths et al., 2002; Decroocq et al., 2003; Eujayl et al., 2004), facilidade de detecção, abundância relativa, cobertura extensiva do genoma (Li et al., 2002) e necessidade de quantidades muito pequenas de DNA para as análises com microssatélites (Powell et al., 1996).

Existem duas principais metodologias utilizadas para o desenvolvimento de novos locos SSRs: a primeira baseia-se no isolamento e sequenciamento de clones contendo possíveis motivos SSRs, seguidos de desenho de *primers* para as regiões flanqueadoras dos motivos. A segunda baseia-se na construção de bibliotecas enriquecidas em sequências de microssatélites, (Edwards et al., 1996), a partir das quais se efetuam a seleção de clones, o sequenciamentos e o desenho dos *primers* adequados.

Com a disponibilização de bancos de dados de sequências expressas (EST–*Expressed*

Sequence Tags) para inúmeras espécies, tornou-se possível o desenvolvimento de SSR funcionais a partir da busca de motivos SSRs nesses bancos de dados. A busca por SSRs em ESTs é uma estratégia rápida e simples para o estudo da porção expressa do genoma, mesmo para os organismos com genomas grandes, complexos e altamente redundantes, como a cana-de-açúcar. Esta estratégia tem a vantagem da obtenção de um marcador diretamente ligado a um gene, pois os ESTs representam uma sequência parcial do cDNA de um gene que foi expresso em um dado tecido num determinado momento (Sterky e Lundeberg, 2000; Pinto et al., 2006).

Em cana-de-açúcar, a análise de 8.678 sequências ESTs revelou aproximadamente 250 SSRs, a maioria composta por repetições perfeitas de trinucleotídeos, sendo os motivos (CCG)n, (CGT)n e (CCT)n os mais comuns (Cordeiro et al., 2001). Os locos selecionados foram avaliados quanto à transferibilidade e ao polimorfismo existente nos gêneros correlatos *Sorghum* e *Erianthus*.

Utilizando-se os EST-SSRs para análise da variabilidade genética em cana-de-açúcar, observou-se o mesmo valordo conteúdo de informação de polimorfismo (PIC–*Polymorphism Information Content*) entre as variedades de cana-de-açúcar (0,23), tendo ele aumentando entre as espécies *S. officinarum* e *S. spontaneum* (0,62), e alcançando o maior valor (0,80) entre indivíduos dos gêneros *Sorghum* e *Erianthus*. Este resultado revela que as variedades de cana-de-açúcar modernas possuem uma estreita base genética. Apesar disso, os marcadores EST-SSRs conseguiram evidenciar algumas diferenças, auxiliando na caracterização da variabilidade genética existente nos bancos de germoplasma (Cordeiro et al., 2001).

O SUCEST (*Brazilian Sugarcane EST Project*) representa o maior e mais completo banco de dados de ESTs de cana-de-açúcar. Esse banco de dados foi construído com financiamento da

FAPESP e envolveu a colaboração entre 200 pesquisadores de diversas universidades e centros de pesquisas do estado de São Paulo associados a alguns pesquisadores de outros estados do Brasil. Foram geradas 237.954 sequências ESTs, agrupadas em 43.141 *clusters*. Baseando-se na estimativa de redundância interna, concluiu-se que a coleção destes transcritos, poderia indicar cerca de 33.000 genes de cana-de-açúcar (Grivet et al., 2001; Grivet et al., 2003; Vettore et al., 2003). Este banco de dados representa uma fonte apropriada á busca de marcadores candidatos (Camargo, 2000) que podem estar intimamente ligados a genes de importância econômica, com destaque àqueles relacionados ao metabolismo de sacarose, resistência a pragas e doenças e à tolerância a condições adversas de clima e de solo (FAPESP, 2001). Uma vez integrados a um mapa genético, os locos identificados pelos marcadores podem ser avaliados para associação com características de importância agronômica, bem como empregados na seleção assistida por marcadores, já que podem estar ligados aos genes responsáveis por características de interesse.

3.2 Single Nucleotide Polymorphism

Single nucleotide polymorphisms (SNPs) são variações na sequência de DNA que ocorrem quando um único nucleotídeo na sequência do genoma é alterado. Esses polimorfismos, juntamente com as deleções e inserções, são responsáveis pela maior parte da variação genética nos organismos (Cho et al., 1999; Rafalsky e Tingey, 2008) e são amplamente distribuídos pelo genoma, sendo mais abundantes em regiões não transcritas e em regiões que flanqueiam os microssatélites (Mogg et al., 2002; Bundock e Henry, 2004).

Por serem abundantes em muitas espécies de plantas como milho (Tenaillon et al., 2001), aveia (Kanazin et al., 2002) e arroz (Yu et al., 2002) entre outros, os SNPs têm sido

muito utilizados como marcadores, criando perspectivas para facilitar o uso no diagnóstico de doenças, seleção assistida por marcadores (SAM), construção de mapas genéticos de alta resolução e caracterização varietal (Batley et al., 2003), sendo considerados o sistema de marcadores mais "atraente" desenvolvido até o momento (Gupta et al., 2001). Além de apresentarem uma distribuição ampla, esses marcadores apresentam facilidades para a genotipagem em alta escala - *high-throughput* (Giancola et al., 2006; Masouleh et al., 2009), sendo que as tecnologias e abordagens desenvolvidas para estudos em humanos (Gibbs et al., 2003; Frazer et al., 2007; Burton et al., 2007) têm facilitado o uso desses marcadores nos mais diversos organismos (Ferguson et al., 2007; Kim e Misra, 2007; Bain et al., 2007).

Diferentes metodologias vêm sendo utilizadas para identificação de SNPs em diferentes espécies (Suliman-Pollatschek et al., 2002). A busca *in silico* dos SNPs em banco de dados de ESTs apresenta baixo custo relativo às metodologias que envolvem sequenciamento. Além disso, nas buscas é possível escolher como alvo sequências homólogas a genes de interesse.

A busca de SNPs em bancos de sequências vem sendo realizada com sucesso em diferentes espécies que possuem grande quantidade de ESTs em bancos de dados (Buetow et al., 1999; Batley et al., 2003; Kota et al., 2003). Devido aos resultados do programa SUCEST, há uma grande disponibilidade de ESTs em cana-de-açúcar, os quais podem ser utilizados como base para busca *in silico* de SNPs em genes de interesse. Poucos trabalhos foram realizados para avaliar o potencial do uso de banco de ESTs para a busca de SNPs em cana-de-açúcar. Eles incluem o uso do banco de ESTs do SUCEST para encontrar SNPs em regiões transcritas e não transcritas do gene Adh (Grivet et al., 2003), e SNPs para mapear alguns genes candidatos e ESTs (Mcintyre et al., 2006). O uso do banco de dados de ESTs de cana-de-açúcar do Plantdb (http://www.plantdb.org/) para encontrar SNPs em ESTs homólogos a 69 genes candidatos

possibilitou a identificação de 1588 possíveis SNPs (média de 23 SNPs por EST) (Cordeiro et al., 2006). Outras metodologias utilizando sequenciamento de nova geração também vêm sendo utilizadas com sucesso na identificação de SNPs em cana-de-açúcar (Bundock et al., 2009).

Diferentes técnicas de genotipagem de SNPs também vêm sendo desenvolvidas, como a utilização de pirosequenciador (Cordeiro et al., 2006) e de espectrômetro de massas (Sequenom MassArray) (Bundock et al., 2009). Devido ao elevado nível de ploidia da cana-de-açúcar, o pirosequenciador não apresenta resolução suficiente para detecção de alelos com frequências baixas, como nas razões 1:9 ou 1:10, sendo mais indicado o uso de espectrometria de massas (Cordeiro et al., 2006; Bundock et al., 2009).

Enquanto grandes coleções de microssatélites estão disponíveis para muitas espécies de plantas, não há uma fonte comparável de bases de dados sobre o polimorfismo de base única em sequências de DNA, especialmente SNPs e indels. A disponibilidade de marcadores abundantes no genoma facilita a construção de mapas de alta resolução, bem como o mapeamento associativo baseado em desequilíbrio de ligação (Rafalski, 2002). Com o avanço das tecnologias de sequenciamento e a maior disponibilidade de bancos de dados de sequências expressas, a identificação e o uso de SNPs tendem a crescer em plantas.

3.2.1 Genotipagem de SNPs por espectrometria de massas

Em organismos poliploides, como a cana-de-açúcar, a frequência dos SNPs em um loco gênico é determinada pelo número de cromossomos que carregam o gene e pelo número dos diferentes alelos e a frequência de cada alelo que possui o loco SNP. Entretanto, pouco se sabe sobre o grau de similaridade entre os cromossomos homeólogos nos grupos de homologia em cana-de-açúcar, bem como sobre a complexidade e diversidade alélica por loco nessa espécie.

Como resultado, qualquer método utilizado para detectar SNPs em um loco particular de cana-de-açúcar deve ser capaz de determinar a existência de polimorfismo em um dado loco SNP e também, o número de cópias dos alelos (frequência) desse loco SNP em diferentes genótipos. Dessa forma, em poliplóides como é o caso da cana-de-açúcar, um SNP será polimórfico entre dois genótipos quando ele diferir quanto à base analisada ou ao diferente número de cópias alélicas presente em cada um dos genótipos.

Como consequência, a detecção de SNPs em alopoliplóides envolve um sistema que permita simultaneamente a identificação do polimorfismo (variação na base) e também no número de cópias alélicas presente nos genótipos em estudo.

A procura pelo aumento no número de análises de SNPs por genótipo e redução de custos tem levado ao rápido desenvolvimento de métodos de genotipagem de SNPs. Nos últimos anos, a espectrometria de massas com ionização por dessorção a laser auxiliada por matriz, e análise por tempo de vôo (MALDI-TOF MS – "*matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*") emergiu como um dos métodos mais poderosos de genotipagem de SNPs em diplóides. A precisão, velocidade da acumulação de dados, possibilidade de realizar múltiplas reações (*multiplex*) em múltiplas amostras simultaneamente, fornecendo resultados da presença e frequência do polimorfismo nas amostras são as principais características de MALDI-TOF MS. Muitos métodos de genotipagem de SNPs com MALDI-TOF MS têm sido implementados com um alto grau de automação e são aplicados em estudos associados a larga-escala, principalmente pelas indústrias farmacêuticas, as quais têm tido interesses crescentes na área de farmacogenômica. A utilização de MALDI TOF MS para essa finalidade tem sido crescente (Tost e Gut, 2003; Gut, 2004). Os métodos de espectrometria de massas têm sido geralmente utilizados para detectar e quantificar com

precisão alelos com frequências muito baixas, como um pequeno percentual na amostra (Oberacher, 2008).

A espectrometria de massas é um método analítico poderoso e versátil que fornece informações valiosas sobre a composição e estrutura das moléculas sob análise, e também, ajuda a esclarecer sobre a quantidade de analitos específicos em misturas (Oberacher, 2008). Esse método fornece uma solução atrativa para genotipagem de SNPs em diplóides e poliplóides, principalmente porque permite a medição direta e rápida de pequenas sequências do DNA, ao invés de ler apenas uma marcação (fluorescente ou radioativa). Os resultados são facilmente recuperados e analisados por um software automatizado (Gut, 2004).

MALDI foi iniciado em 1988 por Karas e Hillenkamp como um método revolucionário para ionização e análise de massas de muitas biomoléculas. Esses pesquisadores descobriram que a irradiação de cristais formados por pequenas moléculas orgânicas adequadas (chamadas de matriz) com um curto pulso de laser causou uma transferência de energia e o processo de dessorção, produzindo íons em fase de gás a partir da matriz. De maior importância ainda, eles descobriram que se uma baixa concentração de um analito não-absorbante, como uma proteína ou molécula de ácido nucléico, fosse adicionada à matriz em solução e embutida nos cristais sólidos da matriz formados por secagem da mistura, as moléculas intactas não-absorbantes do analito também seriam levadas à fase de gás e ionizadas sob irradiação a laser, facilitando sua análise de massas (Griffin e Smith, 2000).

Os componentes básicos de um espectrômetro de massas consistem de uma fonte de ionização (laser UV), um analisador e um detector. Para a análise, a amostra biológica (mistura de ácidos nucléicos) é misturada com um material (matriz), geralmente um ácido orgânico de baixo peso molecular com forte absorção no comprimento de onda do laser (Tost e Gut, 2003).

Cada amostra assim preparada é depositada em um bloco de silicone de um "chip" (que contêm até 384 blocos) para as análises subsequentes. O processo MALDI-TOF é então iniciado por uma dessorção a laser sequencial da mistura analito-matriz presente em cada bloco de silicone (Irwin, 2008). Os subsequentes processos físicos resultam na formação predominante de íons carregados positiva ou negativamente. Esses íons são extraídos com um campo elétrico e separados em função de suas massas moleculares e de suas cargas (Tost e Gut, 2003). As massas dos compostos de ácidos nucléicos são calculadas através do "tempo de voo" (TOF), que reflete o tempo que um composto laser-ionizado e acelerado requer para ser levado através do tubo de voo (1-2m de comprimento) do analisador TOF e alcançar o detector do instrumento. No detector, os compostos ionizados geram um sinal elétrico que fica gravado por um sistema de dados, sendo finalmente convertido em um espectro de massas (Irwin, 2008). A resolução da atual geração de espectrômetros de massa MALDI permite a fácil distinção da substituição de bases nucleotídicas com variação de massa de 1-7 KDa, o que corresponde ao tamanho de DNA de 3-25 nucleotídeos (Tost e Gut, 2003).

Inicialmente, o MALDI-TOF MS era predominantemente aplicado para a análise de proteínas e peptídeos e somente mais recentemente passou a ser usado para a análise de ácidos nucléicos.

A principal vantagem do MALDI-TOF sobre os métodos convencionais de análise de DNA é a velocidade de aquisição de sinal (cerca de 100 microssegundos para um sinal completo) e o fato de fornecer como resultado o peso molecular da molécula, uma propriedade física intrínseca de cada molécula. Isso exclui a utilização de padrões de tamanho. Os métodos convencionais de eletroforese para separação e detecção de DNA são significantemente mais lentos (Tost e Gut, 2003). Finalmente, é possível a automação completa de todos os passos, da

preparação da amostra até a aquisição e processamento de dados, (Ross et al., 1998), dando ao MALDI-TOF MS um grande potencial para aplicações na análise de ácidos nucléicos em larga escala (Griffin e Smith, 2000).

3.2.2 A plataforma SEQUENOM para genotipagem de SNPs

A empresa SEQUENOM lançou o sistema MASSARRAY em 2000 como uma plataforma automatizada de genotipagem para a detecção de variações genéticas (SNPs) no genoma humano. O sistema baseado em espectrometria de massas tem se tornado uma das principais tecnologias para análises de larga-escala e medidas de alta fidelidade de variações na sequência de ácidos nucléicos (Graber et al., 1999; Jurinke et al., 2005).

Desde então, este sistema tem sido aplicado com sucesso em estudos genéticos associativos de larga escala revelando genes responsáveis pela susceptibilidade à doenças (Bansal et al., 2002; Downes et al., 2004; Kammerer et al., 2004, 2005). Além disso, ele também tem sido aplicado a análises quantitativas de expressão gênica, sequenciamento comparativo, haplotipagem e análise dos mecanismos de regulação epigenética através de metilação (Beaulieu e Kosman, 2003; Stanssens et al., 2004; Tang et al., 2004; Honisch et al., 2004; Ehrich et al., 2005).

Os componentes do sistema MASSARRAY incluem: um robô de manipulação líquida para tratamento automatizado da amostra, um espectrômetro de massas MALDI-TOF otimizado para análises de ácido nucléico, juntamente com algoritmos automatizados por um computador para aquisição de dados e análises. Reagentes e protocolos são padronizados e a preparação da amostra é automatizada e simples (Irwin, 2008).

Todas as aplicações do sistema são baseadas na amplificação alelo-específica, em um

volumes de 5 µl em microplaca de 384 poços. Adaptações para microplacas de 96 poços são praticáveis. Os produtos resultantes da reação de amplificação são utilizados para o processamento pós-PCR (MASSEXTEND ou MASSCLEAVE) (Irwin, 2008).

Os produtos pós–PCR são condicionados e dessalinizados *in situ* através da resina de troca iônica e automaticamente transferidos para os "chips" (SpectroCHIP 384), sendo a transferência efetuada individualmente de cada amostra para um bloco de silicone específico do "chip". Dois chips são analisados por vez no analisador MASSARRAY (Irwin, 2008). Os resultados para uma microplaca com 384 amostras são obtidos em menos de 5 horas e automaticamente carregados em um banco de dados que reconhece, analisa e permite a comparação entre eles.

O analisador MASSARRAY, reduziu muito o tempo de análise de SNPs em grandes quantidades de amostras e criou uma ferramenta analítica de relação custo-eficácia extremamente atrativa para um operador não experiente. O instrumento utiliza um MALDI-TOF MS para análises exatas em alta-escala de DNA e RNA. As vantagens são a velocidade de análise, com um espectro de massas sendo gerado em milissegundos e a detecção direta de uma propriedade intrínseca da amostra: sua massa molecular (Irwin, 2008).

Pelo fato das pequenas diferenças de massas poderem ser detectadas nesse sistema, o MASSARRAY pode conclusivamente distinguir genes parálogos concordantes, como no caso de muitos genomas complexos de plantas. Usando essa tecnologia é possível analisar rearranjos gênicos, eventos de duplicação e deleção tão bem quanto o produto da expressão de genes estreitamente relacionados (Irwin, 2008).

As propriedades características que resumem as vantagens do sistema MASSARRAY estar acima das tecnologias convencionais são: precisão, flexibilidade, "throughput", automação e custo-efetivo (Köster, 2002). Por essas razões o MASSARRAY se mostra promissor para se

tornar um método padrão de identificação e genotipagem de plantas, bem como medições de expressão gênica em níveis de baixa transcrição, características estas de grande importância em qualquer projeto genômico funcional de plantas (Irwin, 2008).

4 Mapeamento genético em plantas

Tradicionalmente são utilizadas populações provenientes do cruzamento entre linhagens endogâmicas, obtidas a partir de autofecundações ou retrocruzamentos sucessivos, para a construção de mapas genéticos. Nestes casos os métodos genético-estatísticos já são bem conhecidos e estão implementados em programas amplamente conhecidos, como por exemplo o MAPMAKER/EXP.

Em contrapartida, existem espécies para as quais a obtenção de linhagens endogâmicas é impraticável devido à depressão por endogamia ou mesmo, ao extenso tempo na obtenção e avaliação destas linhagens (espécies perenes). Nesse caso, a solução encontrada para a obtenção de uma população de mapeamento é o cruzamento biparental entre genitores não homozigóticos, resultando em uma população de irmãos completos (Lin et al., 2003) ou populações F_1 segregantes.

4.1 Mapeamento genético em cana-de-açúcar

No caso da cana-de-açúcar, uma estratégia que foi muito utilizada no passado para a construção de mapas genéticos, denominada *duplo pseudo-testcross,* consiste na construção de mapas individuais (um para cada genitor), através da identificação de marcadores polimórficos entre os genitores segregando em dose única (Grattapaglia e Sederoff, 1994; Porceddu et al., 2002; Shepherd et al., 2003; Carlier et al., 2004). Com base nessa abordagem, mapas de

ligação para *S. officinarum* ('LA Purple') e *S. robustum* ('Mol 5829') foram construídos usando marcadores RAPD, RFLP e AFLP em dose única (Guimarães, 1999). Entretanto, tanto do ponto de vista biológico como estatístico, é desejável a integração das informações contidas nesses mapas individuais em um único mapa integrado, possibilitando a visualização dos dados como um todo. Para isso é necessário a utilização de marcadores que estão presentes em ambos parentais, os quais são então utilizados para estabelecer relações de ligação entre os marcadores que estão segregando individualmente em cada genitor (Wu et al., 1992; Barreneche et al., 1998; Garcia et al., 2006; Oliveira et al., 2007).

Com a utilização dessa abordagem para a construção de mapas genéticos integrados foi possível a utilização de marcadores em diferentes tipos de segregação que não eram utilizados na estratégia anterior *(duplo pseudo-testcross)*, aumentando assim a saturação do mapa e a cobertura do genoma. Para espécies poliplóides como a cana-de-açúcar, os marcadores codominantes são úteis para reunir os grupos de co-segregação em seus respectivos grupos de homologia (Da Silva et al., 1993; Grivet et al., 1996). Além disso, outra vantagem apresentada pela utilização de mapas genéticos integrados é a localização de QTLs com maior precisão (Maliepaard et al., 1998).

Diversos trabalhos abordam a construção de mapas genéticos integrados em populações F_1 segregantes (Ritter e Salamini, 1996; Maliepaard et al., 1997; Ridout et al., 1998). Porém um fator limitante da metodologia apresentada está na impossibilidade de caracterizar corretamente as fases de ligação entre marcadores. Wu et al. (2002a) propuseram um método baseado em máxima verossimilhança que permite estimar simultaneamente a ligação e as fases de ligação entre diversos tipos de marcadores moleculares. A disponibilidade de um mapa em que as fases de ligação foram corretamente estimadas é extremamente importante para o mapeamento de

QTLs, pois permite a aplicação de uma abordagem mais robusta na caracterização das fases de ligação entre marcadores. A imprecisão na caracterização das fases de ligação pode levar a sérios vieses para a estimação de posições e efeitos nessas populações (Lin et al., 2003).

Essa abordagem foi utilizada por Garcia et al. (2006) e Oliveira et al. (2007) na construção de mapas genéticos integrados utilizando um cruzamento biparental de cultivares pré-comerciais de cana-de-açúcar (SP80-180 x SP80-4966). Os resultados obtidos pelos autores foram muito superiores em termos de aproveitamento dos marcadores genotipados e robustez do mapa, quando comparados àqueles obtidos pelo emprego do programa JoinMap (que emprega a estratégia *duplo pseudo-testcross)*, indicando a maior eficiência na estimativa de ligação e fases de ligação do método proposto por Wu et al. (2002a).

A abordagem de Wu et al. (2002a) foi implementada no programa OneMap (Margarido et al., 2007), o qual permite a construção de mapas genéticos em progênies de irmãos completos. Em uma versão recentemente liberada (04/2011) na página web deste programa (http://brieger.esalq.usp.br/CRAN/web/packages/onemap/index.html), foi implementada a abordagem multiponto, baseada em cadeias de Markov Ocultas (Lander e Green, 1987; Jiang e Zeng, 1997; Wu et al., 2002a; Butcher et al., 2002).

Apesar das melhorias obtidas na construção de mapas genético-moleculares com o uso dessa nova abordagem de análise em possível que segregam nas proporções 1:2:1 ou 1:1:1:1 para espécies diplóides (Wu et al., 2002a). Isto se deve às características dos marcadores moleculares utilizados até o momento para poliplóies, para os quais não é possível conhecer a dosagem alélica exata de cada loco. Isso se deve há duas razões: uma relacionada à técnica utilizada para a detecção do polimorfismo e outra associada à complexidade genético-estatística para o mapeamento de locos em dosagens múltiplas.

A presença de alopoliploidia em cana-de-açúcar também dificultou a determinação e uso dos locos em dosagem múltipla, e até o momento há poucos trabalhos que tentaram construir mapas genéticos com marcas apresentando diferentes dosagens no genoma (Da Silva et al., 1993, 1995; Ripol et al., 1999). Sendo assim, torna-se difícil a obtenção dos grupos de ligação e a ordenação dos marcadores nesses grupos quando são construídos mapas genético-moleculares para cana-de-açúcar, resultando em mapas pouco saturados, apresentando menor cobertura do genoma.

OBJETIVOS

Geral

- ✓ Construção de um mapa genético molecular funcional para cana-de-açúcar, integrando novos marcadores moleculares funcionais do tipo EST-SSR e SNPs

Específicos

- ✓ Participação no desenvolvimento de marcadores moleculares do tipo EST-SSR e SNP a partir de buscas realizadas no banco de dados de sequências expressas de cana-de-açúcar (SUCEST) - Artigos I, II, III, IV e V;
- ✓ Otimização da metodologia de genotipagem de SNPs utilizando espectrometria de massas – Artigo V;
- ✓ Identificação da segregação 1:2:1 em SNPs presentes em dose única – Artigo V;
- ✓ Construção de um mapa genético molecular integrando os marcadores EST-SSRs e SNPs presentes em dose única, desenvolvidos neste trabalho, para a população F_1 obtida no cruzamento IACSP 95-3018 x IACSP 93-3046 – Artigo V.

ARTIGO I

"Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs"

L.R. Pinto, K.M.Oliveira, T.G. Marconi, A.A.F. Garcia, E.C. Ulian, A.P. de Souza

Publicado na revista

Plant Breeding (125: 378-384; 2006)

Plant Breeding 125, 378—384 (2006)

Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs

L. R. Pinto[1], K. M. Oliveira[2], T. Marconi[2], A. A. F. Garcia[3], E. C. Ulian[4] and A. P. de Souza[2]

[1]Centro Avançado da Pesquisa do Agronegócio de Cana, Instituto Agronômico de Campinas-IAC, Anel Viário Contorno Sul, km 321, CP 206, CEP 14001-970, Ribeirão Preto, São Paulo, Brasil, E-mail: lurossini@iac.sp.gov.br; [2]Centro de Biologia Molecular e Engenharia Genética (CBMEG) – Universidade Estadual de Campinas (UNICAMP), Cidade Universitária 'Zeferino Vaz',CP 6010, CEP 13083-970, Campinas-São Paulo, Brasil; [3]Depto de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba-São Paulo, Brasil; [4]Centro de Tecnologia Canavieira, Piracicaba-São Paulo, Brasil

With 2 figures and 3 tables

Received October 21, 2005/Accepted January 23, 2006
Communicated by W. E. Weber

Abstract

Microsatellites or simple sequence repeats (SSRs) are one of the most suitable markers for genome analysis as they have great potential to aid breeders to develop new improved sugarcane varieties. The development of SSR derived from expressed sequence tags (EST) opens new opportunities for genetic investigations at a functional level. In the present work, the polymorphism obtained with a subset of 51 EST–SSRs derived from SUCEST was compared with those generated by 50 genomic SSRs (gSSR) in terms of number of alleles, polymorphism information content, discrimination power and their ability to establish genetic relationships among 18 sugarcane clones including three *Saccharum* species (*S. officinarum*, *S. barberi*, *S. sinense*). The majority of EST–SSRs loci had four to six alleles in contrast to the seven to nine observed for the gSSRs loci. Approximately, 35% of the gSSRs had PIC values around 0.90 in contrast to 15% of the EST–SSRs. However, the mean discrimination power of the two types of SSR did not differ significantly as much as the average genetic similarity (GS) based on Dice coefficient. The correlation between GS of the two types of SSRs was high ($r = 0.71/P = 0.99$) and significant. Although differences were observed between dendrograms obtained with each SSR type, both were in good agreement with pedigree information. The *S. officinarum* clone IJ76-314 was grouped apart from the other clones evaluated. The results here demonstrate that EST–SSRs can be successfully used for genetic relationship analysis, extending the knowledge of genetic diversity of sugarcane to a functional level.

Key words: *Saccharum officinarum* — microsatellites — EST

Modern sugarcane varieties are mainly derived from interspecific crosses between the noble cane *S. officinarum* (2n = 80) and the wild species *S. spontaneum* (*2n* = 40–128) that was chosen for its resistance to pathogens and other stresses. Successive backcrosses of the hybrids with *S. officinarum* were carried out to recover the high-sugar producing types. Thus, present-day sugarcane cultivars possess a complex interspecific aneu-polyploid genome with chromosome numbers ranging from 100 to 130 (Hoarau et al. 2005). This genomic complexity added to the multigenic and/or multi-allelic nature of most agronomic traits makes sugarcane breeding a difficult task (Casu et al. 2005).

Molecular markers have a great potential to help breeders to develop new improved sugarcane varieties. Microsatellites or simple sequence repeats (SSRs) are one of the most suitable markers for genome analysis. They are short segments of DNA of one to six tandem repeated base pairs that are commonly isolated from enriched libraries in a laborious and time-consuming process (Jarne and Lagoda 1996). Microsatellite markers have many applications in sugarcane genetics and breeding processes including germplasm analysis, cultivar identification, parent evaluation, genetic mapping and marker-assisted selection (Cordeiro et al. 2003, Pan et al. 2003a,b, Aitken et al. 2005). The International Sugarcane Microsatellite Consortium funded by the International Consortium for Sugarcane Biotechnology (ICSB) developed 259 genomic microsatellites (gSSRs) to be used for these purposes (Cordeiro et al. 1999, 2000).

Nowadays microsatellites have been developed for many plant species through SSR mining in expressed sequence tags (EST) obtained from partial sequencing of cDNAs (Cho et al. 2000, Scott et al. 2000, Thiel et al. 2003). A large quantity of these ESTs has been deposited in the EST databases (http://ncbi.nlm.nih.gov/dbEST/). As ESTs represent part of expressed genes, they can provide direct mapping of genes with known functions allowing candidate genes, involved in important metabolic pathways to be directly evaluated for associations with important agronomic traits. Moreover, they are ideal for marker-assisted selection as they themselves may be the gene which is responsible for the trait of interest (Cato et al. 2001, Ma et al. 2004).

The SUCEST, Sugarcane Expressed Sequence Tag Project (http://sucest.lad.ic.unicamp.br/en/) has produced a large collection of ESTs with a great potential to generate molecular markers. By mining dbSUCEST for SSR, 2005 clusters containing SSR (EST–SSRs) that are being used as molecular markers in a sugarcane mapping project (Pinto et al. 2004, Garcia et al. 2006) were found. The development of microsatellites derived from expressed sequence tags (EST) opens new opportunities for genetic investigations at a functional level. In an attempt to explore the large amount of sequence information deposited in the SUCEST database for sugarcane breeding, the Plant Molecular Biology Laboratory (UNICAMP, Campinas, São Paulo, Brazil), in a collaborative effort with the Centro de Tecnologia Canavieira (CTC, Piracicaba, São Paulo, Brazil), is developing a working primer set of EST–SSRs.

The present study was undertaken to evaluate the potential of a novel set of (EST–SSRs) developed from the SUCEST database as molecular markers for sugarcane breeding. The objectives were to (i) characterize the polymorphism levels of these novel EST–SSRs, (ii) verify the discrimination power of the EST–SSRs for fingerprinting, and (iii) compare the polymorphism between EST–SSRs and gSSR considering a common subset of sugarcane clones.

Materials and Methods

Plant materials: The level of polymorphism generated by EST–SSRs and gSSRs was evaluated against 13 commercial clones of sugarcane, *Saccharum officinarum*, three *Saccharum* species (*S. officinarum*, *S. barberi* and *S. sinense*) and the parents (SP80-180 and SP80-4966) of a mapping population from the Centro de Tecnologia Canavieira (CTC-Piracicaba, Sao Paulo, Brazil, Table 1).

DNA extraction and PCR amplification: A modified CTAB-method (Hoisington et al. 1994) was used to extract total genomic DNA from 300 mg of lyophilized young leaf tissue ground to a fine powder. PCR reactions were performed in a 20-µl final volume containing 40 ng of template DNA, 0.2 µM of each forward and reverse primers, 100 µM of each dNTP, 2.0 mM $MgCl_2$, 10 mM Tris-HCl, 50 mM KCl, and 0.5 unit *Taq* DNA polymerase (Invitrogen, São Paulo, Brasil). Reactions were amplified on a PTC-100 thermocycler (MJ Research, Watertown, MA, USA) as follows: 94°C for 3 min; followed by 30 cycles of 94°C for 1 min; annealing temperature specific for each primer for 1 min; extension of 72°C for 1 min and a final elongation step at 72°C for 2 min. Amplification products were separated by electrophoresis on 6% denatured PA gels using a 25-bp ladder as a size standard and silver-stained according to Creste et al. (2002).

SSRs: EST–SSRs were developed from cluster consensus sequences derived from the Sugarcane Expressed Sequence Tag database (SUCEST/http://sucest.lbi.dcc.unicamp.br/en/) as detailed described by Pinto et al. (2004). A set of 100 EST–SSR was designed using the Primer Select software (LaserGene versions 5.01/5.02, DNAStar, Inc., Madison, WI, USA) and synthesized by Invitrogen. The stringency criteria used were the same as described by Cordeiro et al. (2000). The results obtained with EST–SSRs were compared with 50 gSSRs (randomly chosen) developed from the International Consortium for Sugarcane Biotechnology (ICSB; Cordeiro et al. 1999, 2000) screened over the same sugarcane clones and three *Saccharum* species.

Polymorphism analysis: The sugarcane clones were scored for the presence (1) or absence (0) of common bands present in the different patterns. The amplified fragments produced by the SSRs (EST–SSRs and gSSRs) were considered alleles of a single locus. The polymorphic information content for each locus was calculated according to Cordeiro et al. (2003). The discrimination power was calculated for each EST–SSR based on Tessier et al. (1999) and was used as a measure of SSR loci efficiency for variety identification.

Genetic-similarity estimate: The final presence–absence matrix, constructed only with polymorphic bands, was used to estimate the genetic similarity between all the clones evaluated, adopting the Dice similarity coefficient: (Dice 1945). The genetic relationships among the clones were visualized by a dendrogram based on the Unweighted Pair Group Method with Arithmetic Averages, UPGMA (Meyer et al. 2004). The cophenetic coefficient (r_{coph}) was calculated to measure the correspondence between the genetic similarity matrix and the cluster analysis (dendrogram). The above analyses were performed with NTSYS-PC software, version 2.0 (Exeter Software, NY, USA; Rohlf 1993). A bootstrap method with 1000 replicates was performed to verify if the number of polymorphic markers used for genetic similarity estimation was enough to support precise estimates among the sugarcane clones evaluated by the programme (Lima et al. 2001). This was performed using a programme developed in the R software (http://www.R-project.org).

Genetic similarity comparison between EST–SSRs and gSSRs: The two genetic similarity matrices obtained with EST–SSR and gSSRs data were compared with measure the degree of relationship between them by computing the product–moment correlation (r) and the Mantel test statistic (Z) using NTSYS-PC software, version 2.0 (Exeter Software; Rohlf 1993). Such comparison was carried out to verify if the genetic similarities generated by the two types of SSRs data (EST–SSRs and gSSRs) provide similar genetic measures.

Results

Polymorphism analysis

The 100 EST–SSRs were initially tested in agarose gels to verify the presence and quality of product amplification. Out of these, 62 were screened in denaturing polyacrylamide gels for banding patterns of high quality, with 51 EST–SSRs being selected for genetic similarity analysis (Table 2). Most of these EST–SSRs have homology to genes of interest (Table 3). The

Table 1: Sugarcane clones investigated in the present study with their pedigrees relationship covering two generations and their breeding origin

Clone	Pedigree[1]	Origin
CB3624	[POJ2364 × EK28] POJ 2878 × ?	Campos, Brazil
SP791011	[CO419 × CO419] NA5679 × CO775 [POJ2878 × CO371]	São Paulo, Brazil
CB4077	[POJ2364 × EK28] POJ2878 × CO290 [CO221 × D74]	Campos, Brazil
IAC51205	[POJ2364 × EK28] POJ2878 × ?	Campinas, Brazil
RB739359	[CO419 × MZ336] IANE5534 × ?	Republic of Brazil
SP701284	[POJ2878 × ?] B4176 × ?	São Paulo, Brazil
IAC64257	[POJ2878 × CO290] CO419 × IAC49-131 [CP27108 × ?]	Campinas, Brazil
SP711406	[CO419 × CO419] NA5679 × ?	São Paulo, Brazil
SP803280	[CP5530 × CP5376] SP71-10088 × H575028 [H49134 × ?]	São Paulo, Brazil
RB855035	[CP521 × CP48103] L6014 × SP701284 [CB4176 × ?]	Republic of Brazil
SP796134	[H53263 × H507209] H634644 × ?	São Paulo, Brazil
SP792313	[CP5659 × ?] SP71-6106 × ?	São Paulo, Brazil
RB855536	[IAC48-65 × ?] SP70-1143 × RB72454 [CP5376 × ?]	Republic of Brazil
SP49666	[NA5679 × ?] SP711406 × ?	São Paulo, Brazil
SP80180	[B3337 × ?] B3337 × ?	São Paulo, Brazil
Ganda Cheni (*S. barberi*)	*S. barberi* × ?	Saretha, India
IJ76-314 (*S. officinarum*)	*S. officinarum* × ?	Iryan, Java
Maneria (*S. sinense*)	*S. sinense* × ?	Pansahi, China

[1]Extracted from Caiana/IAC software.
?, polycross.

Table 2: Polymorphism characterization for 51 EST–SSRs and 50 gSSRs. Number of alleles (A), expected product size, allele range (bp), polymorphism information content (PIC) and discrimination power (*D*)

EST–SSR[1]	A	Expected size[2]	Allele range (bp)[3]	PIC	*D*	gSSR	A	Expected size[2]	Allele range (bp)[3]	PIC	*D*
SCA03	5	181	187–209	0.71	0.79	SMC863CG	7	296	268–309	0.82	1.00
SCA04	3	170	184–227	0.59	0.21	SMC765BS	8	298	46–326	0.83	0.98
SCA06	9	238	215–233	0.81	0.98	SMC39BUQ	10	143	145–163	0.87	0.97
SCA07	8	201	207–227	0.82	0.96	CIR67	12	142	72–139	0.89	0.98
SCA10	7	195	204–248	0.80	0.99	SMC415MS	4	338	284–304	0.86	0.99
SCA12	11	269	223–251	0.87	0.98	SMC2055FL	10	318	248–272	0.83	0.96
SCA14	7	186	193–207	0.78	0.97	SMC21AS	15	133	121–146	0.91	1.00
SCA16	10	165	167–189	0.87	0.99	SMC119CG	12	119	109–168	0.89	1.00
SCA17	11	212	146–192	0.75	0.82	CIR36	16	166	128–158	0.91	1.00
SCA19	2	191	202–216	0.21	0.30	CIR55	9	333	288–341	0.83	0.99
SCA22	4	185	180–194	0.65	0.80	SMC36BUQ	6	118	117–163	0.76	0.93
SCA26	10	183	171–200	0.87	0.98	SMC319CG	10	183	139–166	0.87	0.93
SCB01	4	193	165–196	0.70	0.91	SMC1055HA	7	224	212–278	0.79	0.99
SCB02	6	164	182–205	0.34	0.81	CIR35	8	173	154–182	0.80	0.98
SCB06	5	176	163–193	0.67	0.70	CIR14	5	240	223–236	0.64	0.79
SCB07	13	291	223–300	0.90	1.00	CIR11	7	235	214–156	0.83	0.56
SCB08	3	176	176–183	0.59	0.45	CIR21	18	181	106–149	0.91	1.00
SCB12	6	147	107–182	0.71	0.60	CIR56	17	184	128–170	0.90	1.00
SCB13	5	133	132–145	0.67	0.86	SMC843BS	12	201	172–208	0.90	0.97
SCB14	12	297	263–289	0.90	1.00	SMC687CS	5	203	195–211	0.67	0.81
SCB16	6	234	262–306	0.71	0.87	SMC1011HA	8	276	174–233	0.83	0.81
SCB17	6	158	94–166	0.80	0.86	CIR58	7	318	287–299	0.90	1.00
SCB25	6	191	192–207	0.78	0.86	CIR12	14	279	151–279	0.81	1.00
SCB27	6	213	203–232	0.78	0.94	SMC1235FL	9	88	76–107	0.83	1.00
SCB35	6	180	199–217	0.74	0.82	SMC448MS	17	187	113–232	0.92	1.00
SCB37	6	187	182–214	0.77	0.87	SMC1047HA	15	166	137–171	0.89	1.00
SCB38	5	210	207–229	0.72	0.81	CIR24	7	265	216–242	0.78	0.93
SCB39	8	146	111–144	0.83	0.96	CIR25	5	300	274–293	0.68	0.90
SCB40	7	155	165–188	0.79	0.91	CIR26	10	134	116–138	0.83	0.92
SCB41	21	163	81–165	0.93	0.98	CIR28	12	419	255–315	0.89	1.00
SCB43	4	330	255–290	0.67	0.78	CIR50	9	263	185–298	0.83	0.97
SCB45	12	110	94–160	0.84	0.99	CIR4	7	289	263–298	0.81	0.98
SCB47	7	236	220–242	0.82	0.83	SMC232MS	3	157	162–210	0.47	0.21
SCB48	3	218	199–213	0.57	0.29	SMC477CG	7	168	88–139	0.79	0.64
SCB49	3	454	322–373	0.59	0.60	SMC31CUC	13	225	149–190	0.89	1.00
SCB51	6	416	339–372	0.76	0.97	SMC2017FL	14	229	215–252	0.90	1.00
SCB52	11	145	89–149	0.87	0.95	SMC1069HA	5	195	155–206	0.55	0.81
SCB53	12	113	95–142	0.90	0.81	SMC222CG	9	198	158–183	0.85	0.97
SCC01	15	298	266–306	0.90	1.00	CIR23	9	281	261–332	0.85	0.95
SCC02	5	202	198–217	0.73	0.28	CIR51	6	272	251–279	0.72	0.75
SCC03	2	227	230–242	0.30	0.35	SMC2039FL	7	275	146–163	0.79	0.91
SCC04	9	218	198–214	0.84	0.98	SMC2024FL	8	275	193–206	0.84	0.97
SCC05	8	179	149–196	0.82	0.97	SMC260MS	5	236	201–223	0.71	0.93
SCC07	4	228	201–228	0.62	0.83	SMC280CS	3	238	215–221	0.41	0.11
SCC10	6	252	220–262	0.69	0.67	CIR32	15	211	162–248	0.90	1.00
SCC13	7	356	303–352	0.79	0.97	SMC236CG	8	210	161–225	0.82	0.53
SCC15	5	203	152–215	0.71	0.84	CIR18	8	215	170–189	0.84	0.96
SCC16	3	229	227–251	0.45	0.31	CIR31	7	215	183–197	0.81	0.94
SCC17	14	137	115–150	0.89	1.00	CIR1	21	227	118–244	0.93	1.00
SCC18	4	129	121–132	0.53	0.65	CIR74	8	226	210–232	0.78	0.75
SCC19	8	238	242–281	0.78	0.94						
Total	369			—	—		474			—	—
Mean	7.2			0.73	0.80		9.5			0.82	0.89
$CI_{(95\%)}$	—			0.69–0.77	0.74–0.86		—			0.79–0.85	0.84–0.94

[1]EST–SSR primer sequences are under a confidentially agreement. Contact A. P. de Souza (anete@unicamp.br).
[2]Expected size of the amplified product obtained by primer select (Laser gene) programme.
[3]Allele range estimated by linear regression.

51 EST–SSRs detected 369 alleles of which 64 (17%) were monomorphic among the clones evaluated. Allele number ranged from 2 (SCA19; SCC03) to 21 (SCB41) with a mean of 7.2 alleles per EST–SSR. Most of the EST–SSRs showed high values of polymorphic information content (PIC) with an average of 0.73. The greatest PIC value was observed for SCB41 (0.93) whereas the lowest one (0.21) was for SCA19.

The 50 gSSRs used for polymorphism comparison generated 474 alleles of which 44 (9%) were monomorphic. The number of alleles varied from 3 (SMC232MS) to 21 (CIR1), with a mean of 9.5 alleles per gSSR. The PIC values ranged from 0.41 (SMC280CS) to 0.93 (CIR1) with a mean value of 0.82.

Overall, the polymorphism obtained with gSSR, in a similar sample of sugarcane clones, was higher than those obtained with EST–SSRs. This was based on the comparative distributions of polymorphism shown in Fig. 1. The majority of the EST–SSR loci had four to six alleles in contrast with the seven to nine observed for gSSR loci. Approximately, 35% of gSSRs

Table 3: Relation of the 51 EST–SSRs with their motif and expected EST homology

EST–SSR[1]	Motif	EST Homology[2]
SCA03	(AC)20	(AC006234) (1–4)-beta-mannan endohydrolase
SCA04	(TG)9	(AL391711) putative protein
SCA06	(TG)12	q42976|nlt4_orysa non-specific lipid-transfer protein 4 precursor
SCA07	(AC)7	S65572 pattern-formation protein GNOM
SCA10	(GA)7	No hits found
SCA12	(TA)29	No hits found
SCA14	(AG)23	No hits found
SCA16	(AG)25	(AB009048) gene_id:K15E6.9 unknown protein
SCA17	(AG)32	(AP002482) ESTs D41739(S4522)
SCA19	(CG)9	(AC006592) putative peroxidase
SCA22	(TA)30	No hits found
SCA26	(TG)11	T05613 hypothetical protein F9D16.
SCB01	(CGG)7	(AF034945) glycine-rich RNA binding protein
SCB02	(AGA)5	(AC082644) putative centromere/microtubule binding protein
SCB06	(GCC)8	(AC018363) putative protein phosphatase-2C (PP2C)
SCB07	(CGA)8	(AB013392) pyrophosphate-dependent phosphofructo-1-kinase-like protein
SCB08	(GGT)6	(AF315811) RNA-binding protein
SCB12	(GCC)5	(AB007649) gene_id:MLE2.11
SCB13	(ACG)5	No hits found
SCB14	(CGT)8	(D28861) RNA binding protein, RZ-1
SCB16	(CGG)5	(AP001168) Similar to mRNA for DREB1A (AB007787
SCB17	(GTC)6	(AP002524) contains ESTs AU032708(S13763), D47951(S13763)
SCB25	(AAT)7	(AC003673) putative protein kinase
SCB27	(AGT)6	(AL163815) putative protein
SCB35	(CCG)5	(X67324) MFS18
SCB37	(GCC)6	(L14063) O-methyltransferase
SCB38	(GTC)8	No hits found
SCB39	(CGC)8	(AB023482) ESTs AU058081(E30812)
SCB40	(GCC)5	(X97022) lamin
SCB41	(CGA)8	(AL049525) putative protein
SCB43	(CGA)9	(AP002743) putative pyrophosphate-dependent phosphofructo-1-kinase
SCB45	(GTA)7	(AC004665) putative phosphomannomutase
SCB47	(CGC)5	(AP002746) putative zinc finger transcription factor
SCB48	(CTA)5	(AC004665) hypothetical protein
SCB49	(GCC)5	(AF310215) glycine-rich RNA-binding protein
SCB51	(TCA)6	(AC008261) hypothetical protein
SCB52	(CCG)5	No hits found
SCB53	(GCG)5	No hits found
SCC01	(GATA)14	(Z97022) cysteine proteinase
SCC02	(CGGC)5	(AF130975) plasma membrane intrinsic protein
SCC03	(CCAC)4	(AF241166) MAP kinase MAPK2
SCC04	(GGAT)6	(AB019235) contains similarity to DNA-binding protein~gene_id:MMI9.9
SCC05	(TGCT)4	(X75670) cytochrome b5
SCC07	(CTAG)3	(AB015475) contains similarity to gibberellin-stimulated transcript 1 like protein~gene_id: MMN10.7
SCC10	(AGGC)3	No hits found
SCC13	(TGCG)5	(AC084320) putative actin-depolymerizing factor
SCC15	(TAAT)13	No hits found
SCC16	(TCTA)4	(M57249) phospholipid transfer protein
SCC17	(ACGC)5	(AF060198) PsbY precursor; putative photosytem II peptide
SCC18	(AGGA)3	(AB009053) permease 1
SCC19	(AGCC)5	No hits found

[1]EST–SSR primer sequences are under a confidentially agreement. Contact A. P. de Souza (anete@unicamp.br).
[2]EST homology: annotation of the best homologue identified by BLASTX (available in SUCEST database).

had PIC values around 0.90, in contrast to 15% of the EST–SSRs.

Discrimination power

The discrimination power (D) was high for the majority of EST–SSRs (with an average value of 0.80 reaching the maximum value (1.00) for EST–SSRs SCB07, SCB14, SCC01 and SCC17 (Table 2). For gSSRs, it ranged from 0.11 (SMC80CS) to 1.0 (SMC863CG, SMC21AS, SMC119CG, CIR36, CIR21, CIR56, CIR58, CIR12, SMC1235FL, SMC448MS, SMC1047HA, CIR28, SMC31CUQ, SMC2017FL, CIR32, CIR1) with a mean value of 0.89. Although parameters such as number of alleles and PIC were less informative for the EST–SSR loci, no differences were found between discrimination power averages for the two types of SSR (Fig. 1). In fact, for this parameter, the mean value obtained by the two types of SSRs did not differ considering the significant confidence interval of 95% (Table 2).

Genetic similarity

Genetic similarities (GS) based on Dice coefficients were obtained for all possible 153 pairwise comparisons calculated with 305 polymorphic markers amplified by 51 EST–SSRs.

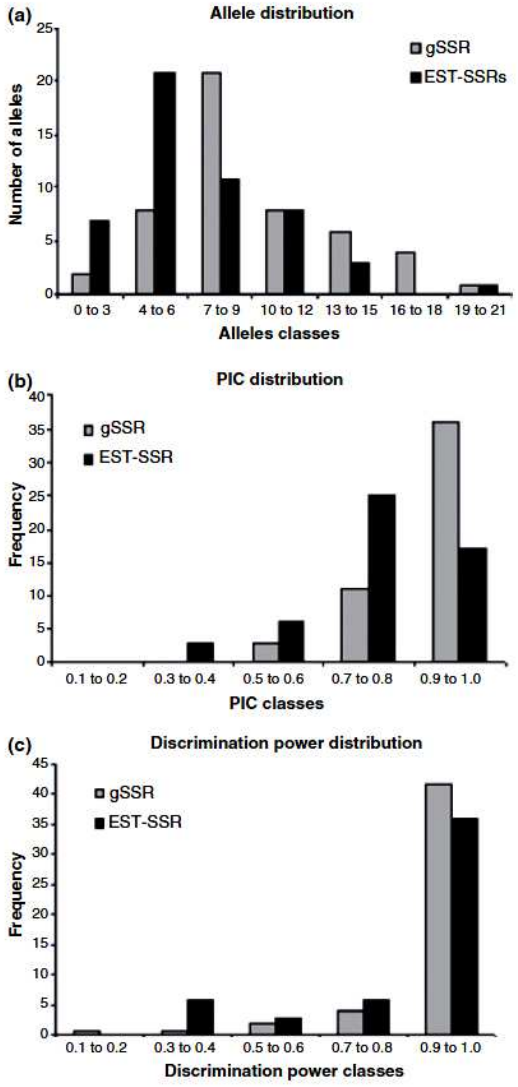


Fig. 1: Comparative distributions of (a) number of alleles, (b) Polymorphism information content (PIC) and (c) discriminatory power (d) between gSSRs and EST–SSRs

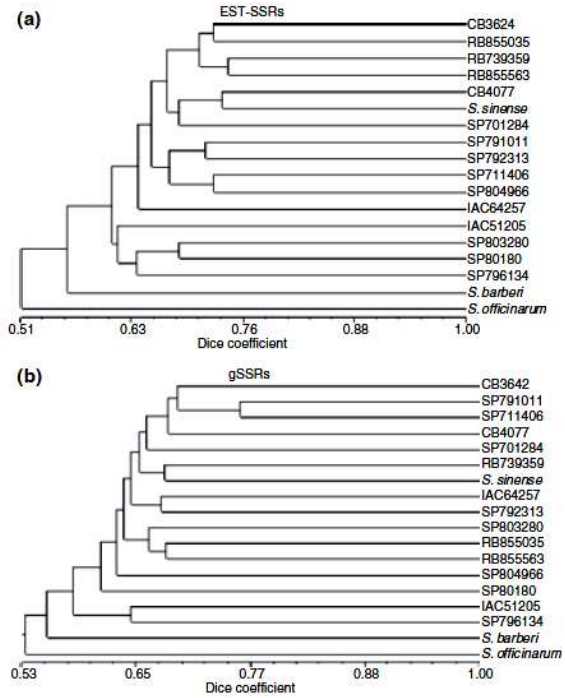


Fig. 2: Dendrogram base on the Dice similarity coefficient and UPGMA clustering method. (a) Calculated with EST–SSRs data. (b) Calculated with gSSRs data

These 153 pairwise combinations gave an average GS value of 0.62. Genetic similarity values varied from 0.46 (SP791011 and IJ76-314) to 0.74 (RB855563 and RB739359). Among the species evaluated, *S. barberi* was more related to *S. sinense* (0.57) than *S. officinarum* (0.52) while *S. barberi* vs. *S. officinarum* was more distant (0.48). The mean GS value among these species was 0.52. A genetic similarity value of 0.69 was found between SP80180 and SP804966. These two sugarcane elite clones were selected from the CTC (Centro de Tecnologia Canavieira, Piracicaba, São Paulo, Brazil) breeding programme and are being used in the construction of a genetic map (Garcia et al. 2006) and for the mapping of QTL.

For gSSR, GS estimates were based on 430 polymorphic markers derived from 50 loci. The highest GS-value (0.75) was observed between the SP791011 and SP711406 clones that are half sibs (Table 1). The most genetically distant clones were *S. officinarum* and IAC51205; *S. officinarum* and SP804966 showed a GS-value of 0.49. On average, the GS-value obtained with gSSR was 0.61. The GS between *S. officinarum* and *S. sinense* was almost the same (0.58) as that observed between *S. barberi* and *S. sinense* (0.57) while *S. barberi* and *S. officinarum* was the most genetically distant (0.51). The mean GS-value among these species was 0.55. For the SP80180 and SP804966 mapping parents, the GS-value was 0.61 and equivalent to the average.

Based on bootstrap analysis (data not shown), the mean coefficient variation (CV) value obtained with 430 gSSR and 305 EST–SSR polymorphic markers was 4.73% and 5.42%, respectively. This implies that the number of polymorphic markers used were sufficient to produce reliable GS estimates. Interestingly, the number of EST–SSR polymorphic markers (80 markers) needed to reach a 10% CV-value that is usually recommended, was lower than that required for gSSRs (90 markers).

Although differences in pairwise GS values were observed for each type of marker (gSSRs and EST–SSRs), the correlation between GS was high ($r = 0.71$/P = 0.99) and significant. This result shows that EST–SSRs can be successfully used for sugarcane genetic relationship analysis.

Cluster analysis

The genetic relationships among the sugarcane clones evaluated for each type of SSR are represented by dendrograms in Fig. 2. The cophenetic values were high and significant for both types of SSR ($r_{EST-SSR} = 0.85$; P = 0.99; $t = 4.4$ and $r_{gSSR} = 0.87$; P = 0.99; $t = 4.6$) indicating a good fit with genetic similarity values. The clones were not separated into clear distinct groups but into a large number of sub-clusters

demonstrating a weak form of structuration among the sugarcane varieties evaluated.

Discussion

In the present work, the polymorphisms obtained with a subset of 50 EST–SSRs were compared with those generated by 50 gSSR, considering parameters such as number of alleles, polymorphism information content, discrimination power and their ability to establish genetic relationships among sugarcane clones.

The results demonstrate that EST–SSR loci are able to generate sufficient polymorphism to ensure their use as molecular markers in sugarcane genetic investigations. The number of alleles and PIC values revealed by the 51 EST–SSRs were close to those reported in sugarcane polymorphism analyses using a few gSSR (Cordeiro et al. 2003, Pan et al. 2003a).

The power of SSRs to distinguish sugarcane varieties have been shown in several studies (Jannoo et al. 2001, Pan et al. 2003b). In fact, the polyploid nature of sugarcane makes SSR markers ideal for fingerprinting (Cordeiro 2001). The lack of difference between discrimination power averages of the two types of SSRs confirms the ability of EST–SSR to generate unique sugarcane fingerprints.

The high value of the average GS (0.61) exhibited with SSRs (EST–SSRs and gSSRs) reveals the narrow genetic base of sugarcane clones, attributed to crosses involving common parents (Table 1). As has been seen, the exchange of materials among breeding programmes has lead to the repeated use of varieties derived from good performing crosses (Nair et al. 2002, Arro 2005) as parents. This is the case of POJ2878, introduced into most sugarcane research stations of the world (Lu et al. 1994), which was used as the progenitor parents of varieties derived from several different breeding programmes (Table 1). Either with gSSRs or EST–SSRs, the highest GS values were observed among related sugarcane clones. This was the case of SP804966 derived from SP711406 and clones SP791011 vs. SP711406, CB3624 vs. CB4077, CB4077 vs. IAC51205 and CB3624 vs. IAC51205 that are half sibs.

The dendrograms generated with each type of SSR did not cluster sugarcane clones into well-defined groups. This result is somewhat expected in sugarcane cluster analysis as a consequence of the high levels of ploidy and heterozygosity which retains a large proportion of alleles derived from the parents involved in the initial interspecific crosses (Lu et al. 1994, Lima et al. 2001) allied to the small number of *S. spontaneum* parents used in the ancestry of most modern varieties. Dendrogram differences observed between EST–SSR and gSSR are probably due to the number of markers and the different portion of genome sampled by each SSR type. Nevertheless, both dendrograms are in agreement with pedigree information covering two generations (Table 1), reinforcing that GS measures derived from SSR markers can be used for planning crosses between divergent sugarcane clones.

Out of the six species that encompass the genus *Saccharum* (*S. officinarum*, *S. barberi*, *S. sinense*, *S. spontaneum* L., *S. robustum*, and *S. edule* Hassk.), three were investigated in this study (*S. officinarum*, *S. barberi*, *S. sinense*) for polymorphism comparisons between EST–SSRs and gSSRs. Because only one accession was used to represent each species, it is difficult to draw well-defined genetic relationships among them. However, for both types of SSR, *S. officinarum* (IJ76-314) formed a single group that was closer to *S. barberi* (Ganda Cheni) and was clearly separated from *S. sinense* (Maneria). This same result was observed using 1121 polymorphic markers derived from AFLP (Lima et al. 2001).

The low levels of polymorphism commonly displayed with molecular markers derived from gene sequences can limit their wide-spread use in genetic analysis. However, compared with other gene-targeted marker systems, the degree of genetic diversity sampled by EST–SSRs seemed promising for generating reliable genetic parameter estimates for assessment of sugarcane variability. As reported by Arro (2005), the cluster patterns generated by TRAP (Target Region Amplification Polymorphism) markers in 63 sugarcane clones did not reflect their pedigree relationships, as half-siblings, and even full-siblings failed to group in the same cluster. On the contrary, as SSRs have high mutation rates and therefore may show high variation between individuals, some degree of polymorphism is more expected, even in conserved regions.

Finally, the fact that most of the EST–SSRs have homology to genes (Table 3) counterbalances their relatively low levels of polymorphism in relation to those derived from anonymous markers such as gSSRs. Thus, EST–SSRs will contribute to the knowledge of genetic diversity of sugarcane, extending it to the functional level.

Acknowledgements

The authors thank Dr Marcos Guimarães de Andrade Landell for supplying pedigree information with the Caiana software. The present research was financed by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 02/01167-1) and CTC (Centro de Tecnologia Canavieira). LRP received postdoctoral fellowship from FAPESP (01/14656) and KMO received a doctorate fellowship from FAPESP (8; 02/00197-4). APS and AAFG received a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Aitken, K. S., P. A. Jackson, and C. L. A. Mcintyre, 2005: A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor. Appl. Genet. **110,** 789—801.

Arro, J., 2005: Genetic Diversity Among Sugarcane Clones Using Target Region Amplification Polymorphism (Trap) Markers and Pedigree Relationships. MS Thesis, Louisiana State University, Louisiana.

Casu, R. S., J. M. Manners, G. D. Bonnett, P. A. Jackson, C. L. McInTyre, R. Dunne, S. C. Chapman, A. L. Rae, and C. P. L. Grof, 2005: Genomics approaches for identification of genes determining important traits in sugarcane. Field Crops Res. **92,** 137—147.

Cato, S. A., R. C. Gardner, J. Kent, and T. E. Richardson, 2001: A rapid PCR-based method for genetically mapping ESTs. Theor. Appl. Genet. **102,** 196—306.

Cho, Y. G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S. R. McCouch, W. D. Park, N. Ayres, and S. Cartinhour, 2000: Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100,** 713—722.

Cordeiro, G. M., 2001: Molecular marker systems for sugarcane germplasm analysis. In: R. J. Henry (ed.), Plant Genotyping – The DNA Fingerprinting of Plants, 129—146. CABI Publishing, Wallingford.

Cordeiro, G. M., T. L. Maguire, R. J. Henry, and K. J. Edwards, 1999: Optimisation of a microsatellite enrichment technique in *Saccharum* spp. Plant Mol. Biol. Rep. **17,** 225—229.

Cordeiro, G. M., G. O. Taylor, and R. J. Henry, 2000: Characterization of microsatellite markers from sugarcane (*Saccharum* spp.) a highly polyploid species. Plant Sci. **155,** 161—168.

Cordeiro, G. M., Y. B. Pan, and R. H. Henry, 2003: Sugarcane microsatellites for assessment of genetic diversity in sugarcane germplasm. Plant Sci. **165,** 181—189.

Creste, S., A. Tulmann Neto, and A. Figueira, 2002: Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gel by silver staining. Plant Mol. Biol. Rep. **19,** 299—306.

Dice, L. R., 1945: Measures of the amount of ecologic association between species. Ecology **26,** 297—302.

Garcia, A. A. F., E. A. Kido, A. N. Meza, H. M. B. Souza, L. R. Pinto, M. M. Pastina, C. S. Leite, J. A. G. da Silva, E. C. Ulian, A. V. O. Figueira, and A. P. Souza, 2006: Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross based on a maximum-likelihood approach. Theor. Appl. Genet. **112,** 298—314.

Hoarau, J. Y., G. M. Souza, A. D'Hont, M. Menossi, L. R. Pinto, A. P. Souza, L. Grivet, C. F. M. Menck, E. C. Vlian, M. Vincentz. Sugarcane, a tropical crop with a highly complex genome. In: J.F. Morot-Gaudry, P. Lea and J.F. Briat. (Org.). Plant genomics. Sciences Publishers, 2006, in press.

Hoisington, D., M. Khairallah, and D. González-De-León, 1994: Laboratory Protocols: CIMMYT. Applied Molecular Genetics Laboratory, Mexico, DF.

Jannoo, N., L. Forget, and A. Dookun, 2001: Contribution of microsatellites to sugarcane breeding program in Mauritius. Proc. Intl Soc. Sugar Cane Technol. **24,** 637—639.

Jarne, P., and P. J. L. Lagoda, 1996: Microsatellites, from molecules to populations and back. Trends Mol. Ecol. **11,** 424—429.

Lima, M. L. A., A. A. F. Garcia, K. M. Oliveira, S. Matsuoka, H. Arizono, C. L. Jr de Souza, and A. P. de Souza, 2001: Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor. Appl. Genet. **104,** 30—38.

Lu, Y. H., A. D'Hont, F. Paulet, L. Grivet, M. Arnaud, and J. C. Glaszmann, 1994: Molecular diversity and genome structure in modern sugarcane varieties. Euphytica **78,** 217—226.

Ma, H. M., S. Schulze, S. Lee, M. Yang, E. Mirkov, J. Irvine, P. Moore, and A. Paterson, 2004: An EST survey of the sugarcane transcriptone. Theor. Appl. Genet. **108,** 851—863.

Meyer, A. S., A. A. F., Garcia, and A. P. Souza, 2004: Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L). Genet. Mol. Biol. **27,** 83—91.

Nair, N. V., A. Selvi, T. V. Sreenivasan, and K. N. Pushpalatha, 2002: Molecular diversity in Indian sugarcane cultivars as revealed by randomly amplified DNA polymorphisms. Euphytica **127,** 219—225.

Pan, Y. B., G. M. Cordeiro, E. P. Jr Richard, and R. J. Henry, 2003a: Molecular genotyping of sugarcane clones with microsatellite DNA markers. Maydica **48,** 319—329.

Pan, Y., J. D. Miller, R. J. Schnell, E. P. Richard Jr, and Q. Wei, 2003: Application of Microsatellite and RAPD fingerprints in Florida Sugarcane Variety Program [abstract] . Plant and Animal Genome Abstract XI Abstract Book **W189,** p.43.

Pinto, L. R., K. M. Oliveira, E. C. Ulian, A. A. Garcia, and A. P. Souza, 2004: Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome **47,** 795—804.

Rohlf, F. J., 1993: NTSYS-pc Numerical Taxonomy and Multivariate Analysis System, Version 2.0. Exeter Publications, New York.

Scott, K. D., P. Eggler, G. Seaton, M. Rossetto, E. M. Ablett, L. S. Lee, and R. J. Henry, 2000: Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100,** 723—726.

Tessier, C., J. David, P. This, J. M. Boursiquot, and A. Charrier, 1999: Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor. Appl. Genet. **98,** 171—177.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner, 2003: Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **106,** 411—422.

ARTIGO II

"Funcional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross"

K.M. Oliveira, L.R. Pinto, T.G. Marconi, G.R.A. Margarido, M.M. Pastina, L.H.M. Teixeira, A. Figueira, E.C. Ulian, A.A.F. Garcia, A.P. Souza

Publicado na revista

Plant Breeding (20: 189-208; 2007)

Mol Breeding (2007) 20:189–208
DOI 10.1007/s11032-007-9082-1

Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross

Karine M. Oliveira · Luciana R. Pinto · Thiago G. Marconi · Gabriel R. A. Margarido · Maria Marta Pastina · Laura Helena M. Teixeira · Antônio V. Figueira · Eugênio César Ulian · Antônio Augusto F. Garcia · Anete Pereira Souza

Received: 31 July 2006 / Accepted: 1 February 2007 / Published online: 6 April 2007

Abstract The growing availability of ESTs provides a potentially valuable source of new DNA markers. The authors examined the SUCEST database and developed EST-derived markers. Thus to enhance the resolution of an existing linkage map and to identify putative functional polymorphic gene loci in a sugarcane commercial cross, 149 EST-SSRs and 10 EST-RFLPs were screened in the SP80-180 × SP80-4966 mapping population. With the markers already analyzed in the previous map, 2303 polymorphic markers were generated, of which 1669 (72.5%) were single-dose (SD) markers. Out of these 1669 SD markers, 664 (40%) were scattered onto 192 co-segregation groups (CGs) with a total estimated length of 6.261,1 cM. Using both genomic and EST-derived SSR and RFLP markers, 120 out of the 192 CGs were formed into fourteen putative homology groups (HGs). The EST-derived markers were subjected to BLASTX search in the SUCEST database, of which putative function was assigned to 113 EST-SSRs and six EST-RFLPs based on high nucleotide homology to previously studied genes. The integration of EST-derived markers improved the map, making it possible to consider additional fine mapping of the genome, and providing the means for developing 'perfect markers' associated with key QTL. To summarize, this paper deals with the construction of a genetic linkage map of sugarcane that is populated by functionally associated markers.

Keywords ESTs · Functional map · Mapping · Polyploid · SUCEST · Sugarcane

K. M. Oliveira · T. G. Marconi · L. H. M. Teixeira · A. P. Souza (✉)
Centro de Biologia Molecular e Engenharia Genética (CBMEG), Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, CEP 13083-875 Campinas, SP, Brasil
e-mail: anete@unicamp.br

L. R. Pinto
Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana - IAC/Apta, Anel Viário Contorno Sul, Km 321, CEP 14.001-970CP 206, Ribeirao Preto, SP, Brasil

G. R. A. Margarido · M. M. Pastina · A. A. F. Garcia
Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970 Piracicaba, SP, Brasil

A. V. Figueira
Centro de Energia Nuclear na Agricultura (CENA), Universidade de São Paulo, Avenida Centenário, 303, CP 96, SP 13400-970 Piracicaba, SP, Brasil

E. C. Ulian
Centro de Tecnologia Canavieira - CTC, Caixa Postal 162, 13400-970 Piracicaba, SP, Brasil

Introduction

Modern sugarcane cultivars have a highly polyploid, aneuploid genome with chromosome number ranging from 100 to 130 (Roach 1969). These modern cultivars are interspecific hybrids derived essentially from crosses between *Saccharum officinarum* ($2n = 80$, $x = 10$), a species that has stalks with high sugar content, and *Saccharum spontaneum* ($2n = 40–128$, $x = 8$), a wild, vigorous species that is resistant to several sugarcane diseases. In order to reduce the contribution of *S. spontaneum* and to enhance the sugar content of the hybrids, a series of backcrosses to *S. officinarum* was carried out in early breeding programs, in a process known as nobilization. Hence, modern cultivar genomes are predominantly derived from *S. officinarum*, with only 15–25% of their chromosomes contributed by *S. spontaneum*, as has been shown using genomic in situ hybridization (GISH) (D'Hont et al. 1996) and by molecular marker approaches (Grivet et al. 1996). Because of their high polyploidy and interspecific origin, these cultivars produce progenies with aneuploid chromosome numbers.

Due to the complexity of the sugarcane genome, which appears hard to be deciphered, sugarcane genetics had received little attention from plant scientists. However, a vast array of genomic tools has become available and new opportunities have been opened to refine our understanding of the genetic architecture of sugarcane and to explore its functional system (Hoarau et al. 2006). The development of ubiquitous genetic marker technologies in conjunction with increased computational capability has provided abundant resources for whole-genome linkage analysis.

Among the DNA marker systems available, Microsatellite or Simple Sequence Repeat (SSR) is the most widely used technology for plant genome analysis (Morgante and Olivieri 1993). These markers are characterized by their simplicity, abundance, variation, co-dominance and multi-alleles among genomes. Recent studies revealed that gene transcripts can also contain SSRs, and the abundance of Expressed Sequence Tags (ESTs) has become an attractive potential source of microsatellite markers (Kantley et al. 2002). Indeed, electronic research in genome databases may represent a simple, fast and economical way to obtain SSRs. With the spread of EST projects, SSRs have been developed for many plant species through SSR mining in EST sequences, including *Vitis vinifera* (Decroocq et al. 2003), *Triticum aestivum* L. (Nicot et al. 2004), *Grossypium* (Han et al. 2006) and *Citrus* (Chen et al. 2006). In the past few years, several projects for the sequencing of sugarcane ESTs have been initiated in South Africa (Carson et al. 2000), Brazil (http://sucest.lad.ic.unicamp.br/en/) and Australia (Casu et al. 2001), allowing the development of EST-SSRs for *Saccharum* (Cordeiro et al. 2001; Pinto et al. 2004). The EST collections provide the opportunity for intensive development of functionally associated markers. ESTs can provide direct mapping of genes with known function allowing candidate genes involved in important pathways to be directly evaluated for associations with important agronomic traits (Cato et al. 2001). Moreover, their presence in conserved transcribed regions makes them more transferable across species and to closely related genera, an aspect that would increase their value in a breeding programme.

The Sugarcane Expressed Sequence Tag Project (SUCEST) produced 237.954 ESTs, from 26 cDNA libraries, constructed from several organs and tissues sampled at different developmental stages, assembled in 43.000 clusters (Vettore et al. 2001), giving rise to the most complete EST database for sugarcane (http://sucest.lad.ic.unicamp.br/en/). These 43.000 transcript products are believed to represent 90% of the sugarcane genes with over 30% corresponding to new genes (Vettore et al. 2003). Pinto et al. (2004) mined EST-SSRs in this bank and identified 2005 clusters containing SSRs that in turn represents around 5% of the entire cluster population of the bank.

The first-generation molecular marker-based genetic maps for agronomically important plant species have been largely based on anonymous genetic markers. Due to the fact that polyploidy dictates particular constraints for mapping, the development of a low-density genetic map for sugarcane requires much more work than for a diploid species. In polyploids, there are more than

two alleles, with individual alleles being present in varying number. Thus, to be able to detect a linkage between two loci, it is necessary for both loci to produce a single dose restriction fragment, i.e., a fragment produced by only one copy among all the alleles. When polyploidy is high and pairing is polysomic or irregular, alleles presented as a single copy are much more informative for genetic map construction than any other (Wu et al. 1992). Based on this method, partial genetic maps have been produced for *S. spontaneum* (da Silva et al. 1995), *S. officinarum* (Guimarães et al. 1999) and modern cultivars (Grivet et al. 1996; Aitken et al. 2005; Raboin et al. 2006; Garcia et al. 2006). However, coverage of the genomes surveyed is still incomplete and none of the published maps of sugarcane are saturated up to the present moment. Although the current set of genetic markers provides the means to anchor maps across different pedigrees and to establish linkage with QTL (Quantitative Trait Loci) for agronomic traits, they are not in general saturated enough to narrow the search to small areas in the genome.

The aim of this paper is to report the development of a functionally associated marker-based genetic linkage map for sugarcane, using EST-RFLP and EST-SSR marker technologies. These EST-derived markers were integrated into an existing framework of a genetic map for a commercial cross (Garcia et al. 2006), increasing the marker density and coverage of the existing map. This gene-based map will enable the precision of QTL mapping and would be useful for sugarcane breeding programs.

Materials and methods

Plant material

Genetic mapping was conducted using a population made of 100 random chosen individuals obtained by the cross between the SP80-180 [B3337 × polycross] and SP80-4966 [SP71-1406 × polycross] commercial cultivars. The SP80-180 commercial cultivar, which is the female parent, has lower sucrose content and high stalk production, whereas SP80-4966, the male parent, has higher sucrose and lower stalk production. Both parents and population were developed at the Experimental Station of the Centro de Tecnologia Canavieira-CTC (Camamu-BA, Brazil). Total genomic DNA of mapping progenies was extracted from 300 mg of powdery lyophilized young leaf tissue according to the method described by Hoisington et al. (1994), with minor modifications.

Genotypic data and marker notation

Five types of markers were used to genotype the parents and the 100 progeny individuals. Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP) and SSR markers had already been generated and coded as described in detail by Garcia et al. (2006). EST-SSR and EST-RFLP markers were generated as follows.

EST-SSR

A total of 372 EST-SSRs were developed from the SUCESTdb and screened against the parents and 6 F1 progeny individuals from the F1 mapping population (SP80-180 × SP80-4966) in order to determine polymorphism levels. This includes the EST-SSRs already presented in Pinto et al. (2004, 2006) and also new ones, developed as described by Pinto et al. (2004) (data not shown). From this total, 149 polymorphic loci, with at least two band differences and which segregated in the population sample as dominant genetic markers, were selected and scored across the population. Part of these EST-SSRs were screened in the mapping population according to Pinto et al. (2004). Amplification products were separated by electrophoresis on 6% denatured polyacrilamide gels (*w*/*v*) using a 25 bp ladder as a size standard and then silver stained according to Creste et al. (2001). The other part of these EST-SSRs was screened following the protocol described in Raboin et al. (2006). Then, the samples were resolved on 5% denatured polyacrilamide gels (*w*/*v*), dried and exposed to X-ray film (Fuji RX) for 4 days. In these cases, a 10 bp ladder was used as a size standard.

The nomenclature of these markers was: EST followed by a letter and a two digit code number that represent the motif type and EST-SSR number, respectively (A: for di-; B: for tri-; C: for tetranucleotide primer pairs); followed by a number referring to the amplified allele and a letter to denote parental polymorphism origin. Parental polymorphism origin was designated according to the cross type of marker locus, following notation as per Wu et al. (2002) and detailed in the previous map of this population (Garcia et al. 2006). "D1" corresponds to marker locus that is heterozygous in SP80-180 and homozygous in SP80-4966, "D2" represents marker locus that is heterozygous in SP80-4966 and homozygous in SP80-180, while "C" indicates marker locus that is heterozygous in both parents.

EST-RFLP

Twenty microgrammes of genomic DNA were individually digested with 10 restriction enzymes (*Bam*HI, *Bgl*II, *Dra*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Pst*I, *Sal*I, *Sst*I and *Xba*I) in order to prepare the membranes that were used with EST probes. Restriction fragments were loaded on 0.8% (*w*/*v*) agarose gels that were submitted to electrophoresis in a TAE buffer (40 mM Tris acetate, pH 8.0, 2 mM EDTA) at 25 V for 22 h and transferred to nylon membranes (Hybond-N^+, Amersham, UK). Forty EST probes deriving from SUCEST, obtained from the BCC Center (Brazilian Core Collection Center, Jaboticabal, SP, Brazil), were radioactively labeled with α-dCTP^{32}. Hybridizations were performed in a HYB solution (0.5 M Na_2PO_4 pH 7.2, 1% BSA, 7% SDS, 100 µg/ml sheared herring sperm DNA) at 65°C for 18–24 h. The membranes were washed once for 20 min at 65°C in each of the following solutions: solution I (2 × SSC; 5%, SDS), solution II (1 × SSC; 5% SDS 5%) and solution III (0.5 × SSC; 5% SDS) and placed on an X-Omat (Kodak) film for at least 7 days at –80°C.

Probes from expressed sequences (EST-RFLP) evaluated in the population were named Est plus a number according to the corresponded gene. Markers from the same EST probe were identified by probe code followed by a letter indicating a decrease in the size of the fragment and a letter to denote parental polymorphism origin (D1, D2, C).

Marker scoring

All segregating bands that were distinct and unambiguous were scored independently as a dominant marker, based on the presence (1) or absence (0) in the progeny of the mapping population. Since sugarcane is highly polyploid, only single-dose markers (Wu et al. 1992) were used for map analysis. Each marker was tested against the expected ratios using chi-square test (χ^2) for single dose markers, considering 1:1 and 3:1 ratios. Thus, marker segregation types were identified for deviation from 1:1 and 3:1 segregation ratios, expected for single-dose markers in only one of the parents and single-dose markers in both parents (referred to here as double single-dose markers), respectively. All loci with strong deviation from expected proportions were discarded after controlling type I error for multiple tests (Garcia et al. 2006). The genotypic data from EST-SSR and EST-RFLP single-dose markers were then added to data from AFLP, SSR and RFLP markers used to construct the previous genetic map developed by Garcia et al. (2006), which was used as a framework.

Linkage analysis and map construction

The analysis carried out herewith, basically followed the procedures proposed by Wu et al. (2002) and was implemented in the OneMap software (Garcia et al. 2006), using the previous map as a framework. In short, the two-point analysis was carried out using a minimum LOD Score threshold of 5 and 35 cM (Kosambi 1944) for the recombination fraction. The clusters of linked markers (CGs) were then ordered using the Rapid Chain Delineation algorithm (Doerge 1996) and the final position of markers was refined using three-point analysis based on the likelihood. Markers that could not be correctly positioned were set as accessories. Then, the ordered CGs were assembled into homology groups (HGs) based on common RFLP, SSR, EST-RFLP or EST-SSR markers linked in the coupling or repulsion phase.

Results

EST-derived markers evaluation

EST-SSR

The 149 EST-SSRs produced 576 polymorphic bands of which 332 (57.6%) were single dose markers (1:1) (Table 1). Out of these 332 markers, 161 (48.5%) segregated in the SP80-180 gametes (D1—cross type "*ao* × *oo*") and 171 (51.5%) in the SP80-4966 gametes (D2—cross type "*oo* × *ao*"). An additional 183 (31.8%) EST-SSR double single-dose markers (C—cross type "*ao* × *ao*") were added to linkage analysis (Table 2). The number of segregating markers identified with each EST-SSR ranged from 1 to 10 with an average of 3.5. Neither the markers with two or higher allele dosage nor the markers that did not fit expected segregation ratios (1:1 or 3:1) were exploited in this work.

EST-RFLP

From the 40 EST probes evaluated, 35 (87.5%) were polymorphic between the parents and only 5 (12.5%) were monomorphic. From these 35 EST probe/enzyme combinations, 10 were selected and screened in the mapping population. A total of 47 polymorphic markers were generated, of which 37 (78.7%) were segregating in expected ratios (1:1 and 3:1) (Table 1). Out of these 37 markers, 13 (35.1%) were polymorphic in SP80-180 (D1—cross type "*ao* × *oo*"), 14 (37.8%) were polymorphic in SP80-4966 (D2—cross type "*oo* × *ao*") and 10 (27.1%) were polymorphic in both parental genotypes (C—cross type "*ao* × *ao*") (Table 2). The 10 remaining markers had distorted segregation and were therefore discarded.

Integration of EST-derived markers into the previous map

The 515 EST-SSR and 37 EST-RFLP markers were integrated into a framework map of the SP80-180 × SP8049-66 population (Garcia et al. 2006), in combination with the already mapped AFLP, RFLP and SSR data. Hence, a total of 1669 segregating bands were scored and classified as being from SP80-180 (D1 type) or SP80-4966 (D2 type), or as being a common marker (present in both parents) (C type). Polymorphic and segregating markers included 719 AFLP markers generated by 23 primer pair combinations, 210 SSR markers generated by 52 primer pairs, 188 RFLP markers generated by 55 probe/enzyme combinations and the EST-derived markers (Table 2).

In order to avoid false positives results when testing for linkage, CGs were built up at a stringent LOD score threshold of 5. This high value was chosen due to the high number of markers and the high number of expected CGs, which is supposed to be over 100 CGs because of the high number of chromosomes in modern sugarcane cultivars that typically contain more than 100 chromosomes. As a result, 664 (40%) markers formed 192 CGs. Out of these 664 markers, 179 (27%) were polymorphic on

Table 1 Overall markers screened across progeny of SP80-180 × SP80-4966 cross: number of polymorphic markers and single- and double single-dose markers

Markers	AFLP	RFLP		SSR		
		Genomic	EST	Genomic	EST	Total
Number of markers evaluated	1104	221	47	355	576	2303
Number of polymorphic markers between parents	304	112	35	190	385	1026
Number of monomorphic markers between parents	800	109	12	165	191	1277
Single-dose markers (1:1)	212	100	27	129	332	800
Double-single-dose markers (3:1)	507	88	10	81	183	869
Total number (1:1 and 3:1)	719	188	37	210	515	1669
Number of markers with distorted segregation	385	33	10	145	61	634

Modified from Garcia et al. (2006)

Table 2 Distribution of the different marker types according to their cross type

Cross type	No. of markers					
	RFLP		SSR			
	Genomic	EST	Genomic	EST-SSR	AFLP	Total
D1 (*ao* × *oo*)	47	13	55	161	80	356 (21.3)
D2 (*oo* × *ao*)	53	14	74	171	132	444 (26.6)
C (*ao* × *ao*)	88	10	81	183	507	869 (52.1)
Total	188	37	210	515	719	1669
Cross type	No. of linked markers					
	RFLP		SSR			
	Genomic[a]	EST[a]	Genomic[a]	EST[a]	AFLP[a]	Total[b]
D1 (*ao* × *oo*)	19 (1.1)	5 (0.3)	19 (1.1)	97 (5.8)	39 (2.3)	179 (27.0)
D2 (*oo* × *ao*)	16 (1.0)	4 (0.2)	30 (1.8)	102 (6.1)	76 (4.6)	228 (34.4)
C (*ao* × *ao*)	24 (1.4)	1 (0.1)	24 (1.4)	76 (4.6)	132 (7.8)	257 (38.6)
Total	59	10	73	275	247	664

Modified from Garcia et al. (2006)

[a] Percentage of linked markers in relation to the total number of markers available for mapping (1,669)

[b] Percentage in relation of the total number of linked markers (664)

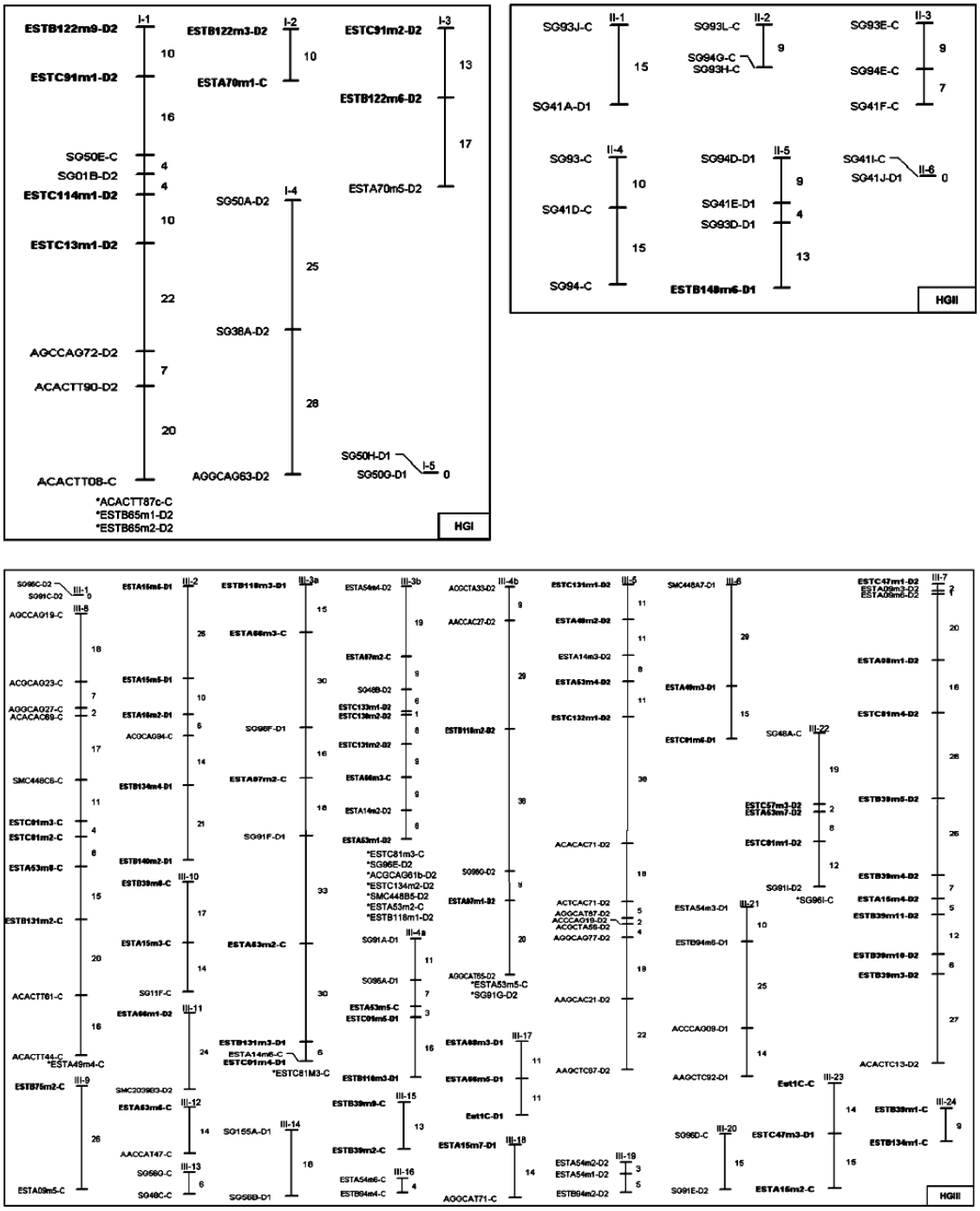
SP80-180 (D1) and linked into the map resulting in 62 CGs, 228 (34.3%) were polymorphic on SP80-4966 (D2) and linked, forming 78 CGs. The remaining 257 (38.7%) linked markers were common from both parents (C) and were used as bridges for combining the information of markers (D1 and D2) into a single map (Table 2). The CGs varied in length from 0 cM to 149.0 cM, with a cumulative length of 6.261,1 cM and an average length of 35.2 cM, with mean distance between markers of 9.4 cM. An irregular distribution of markers along the chromosomes was observed, with an average of 3.6 markers per linkage group (Fig. 1). The 275 EST-SSR and 10 EST-RFLP markers were incorporated into 28 CGs of the previously constructed genetic map and formed 79 new CGs, corresponding to 43% of the linked markers. From these new groups, 39 (49.3%) assembled only EST-derived markers. The remaining 85 CGs were already present in the previous map.

Assignment of HGs using SSR and RFLP markers

Homologous linkage groups were identified using common allelic bridges, provided by SSR and RFLP markers, both genomic and EST-based, due to their co-dominance inheritance. Out of the 192 CGs, 120 could be regrouped into 14 putative HGs, each containing between 2 and 24 CGs. Markers were not evenly distributed over the different HGs, with the largest HG containing 115 markers and the smallest containing 5 (Fig. 1). Only three (HGIV, HGVII, HGXI) of these 14 HGs obtained did not encompass EST-derived markers. Fifty-four CGs containing two mapped markers and 18 CGs containing more than two markers became stranded as independent groups because of the lack of common SSR and RFLP markers to discriminate HGs (Fig. 1). In general, the order of SSR and RFLP markers was consis-

Fig. 1 Functional integrated sugarcane map constructed using 100 individuals obtained from a commercial cross SP80-180 × SP80-4966. Six hundred sixty-four segregating markers were assembled into 192 co-segregation groups (CGs) and 120 of these groups were arranged by 14 homology groups (HGs). The remaining 72 unassigned CGs are numbered consecutively. Positions of loci are given in centiMorgans (Kosambi 1944), on the right of each CG and marker names are on the left. Accessory markers for some CGs are shown with asterisks and below their respective groups. These markers belong to the CG but ordering was not possible. EST-derived markers showing homology to known genes are presented in *bold*

Mol Breeding (2007) 20:189–208



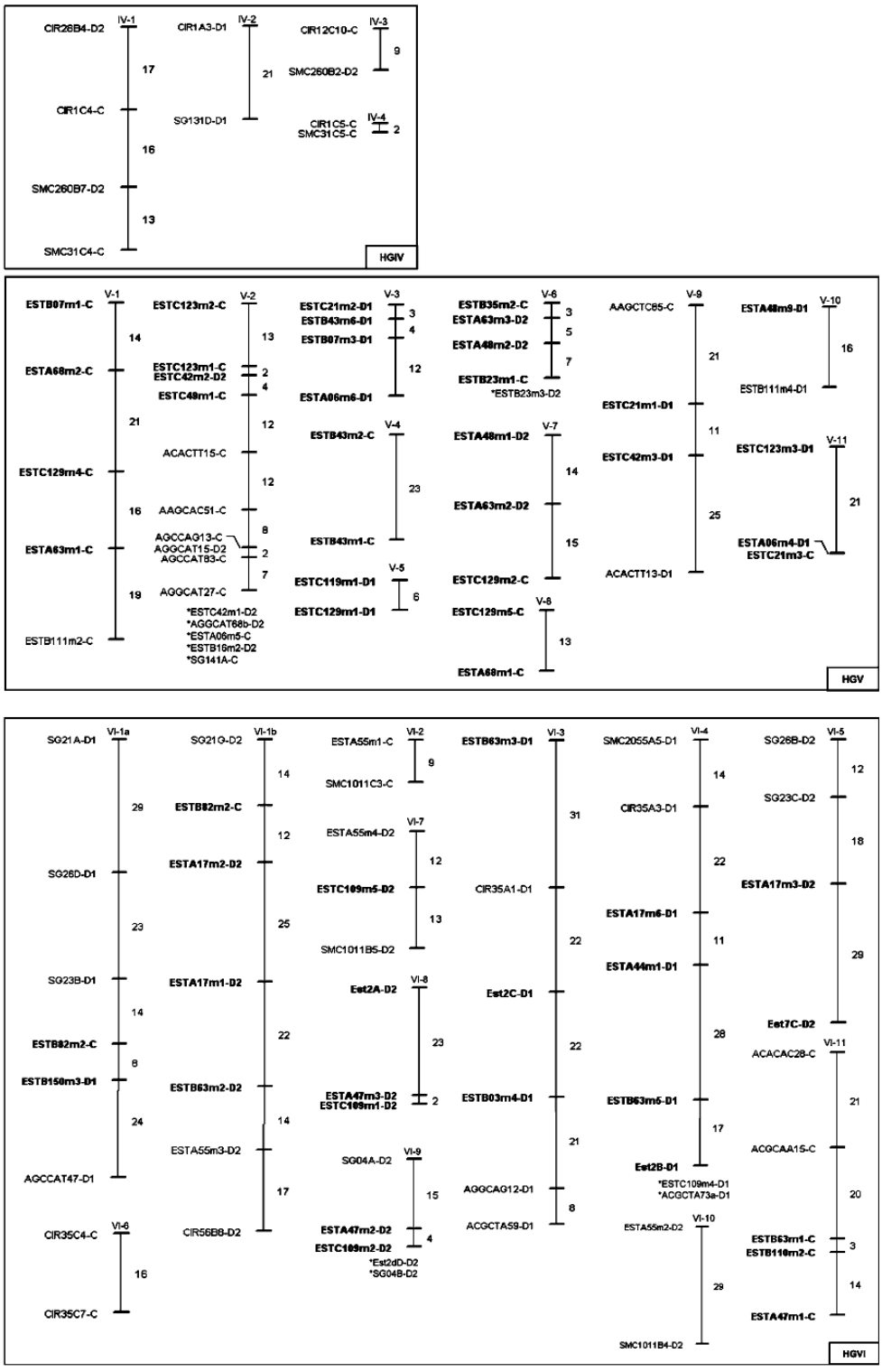


Fig. 1 continued

Mol Breeding (2007) 20:189–208

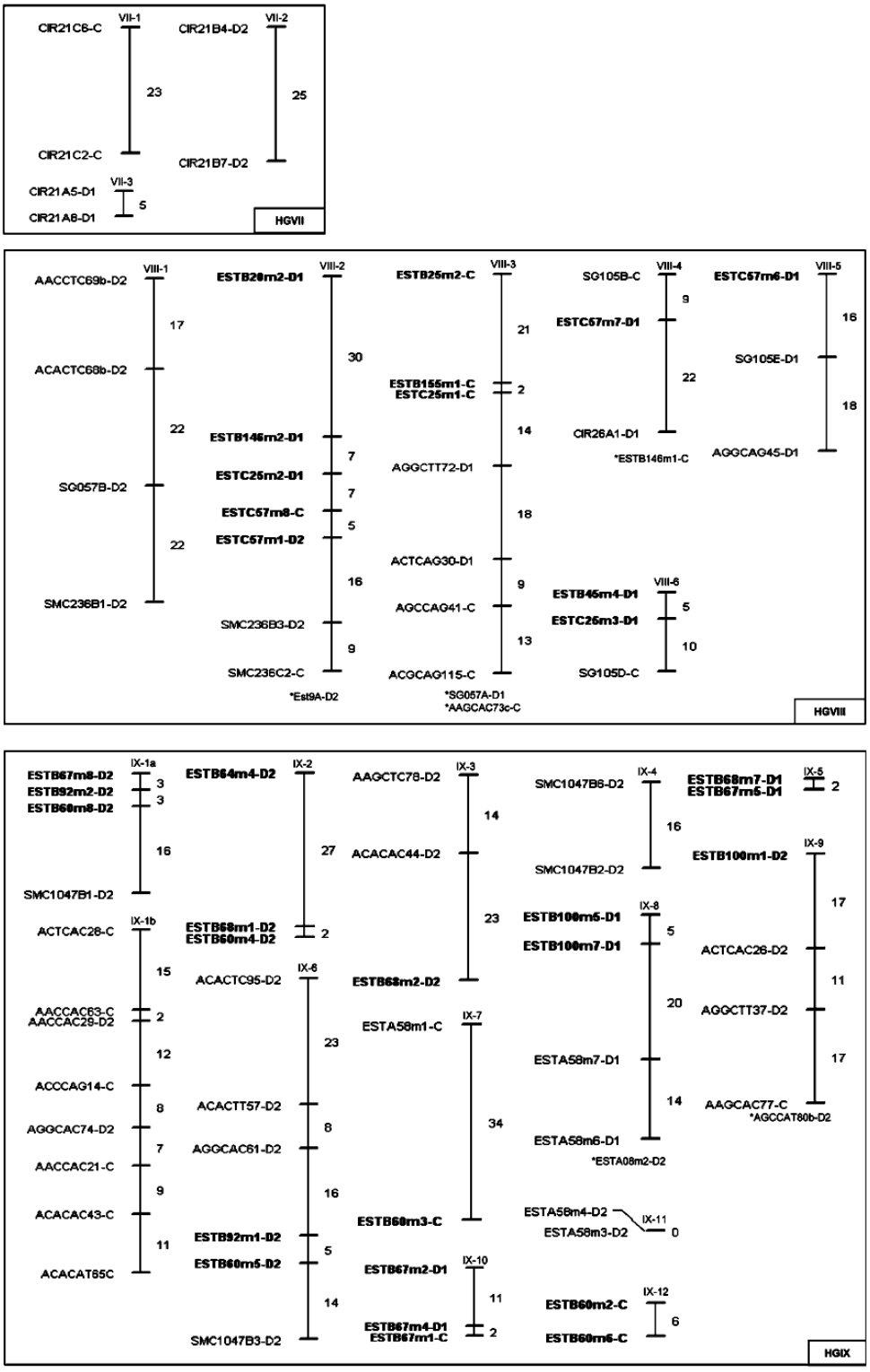


Fig. 1 continued

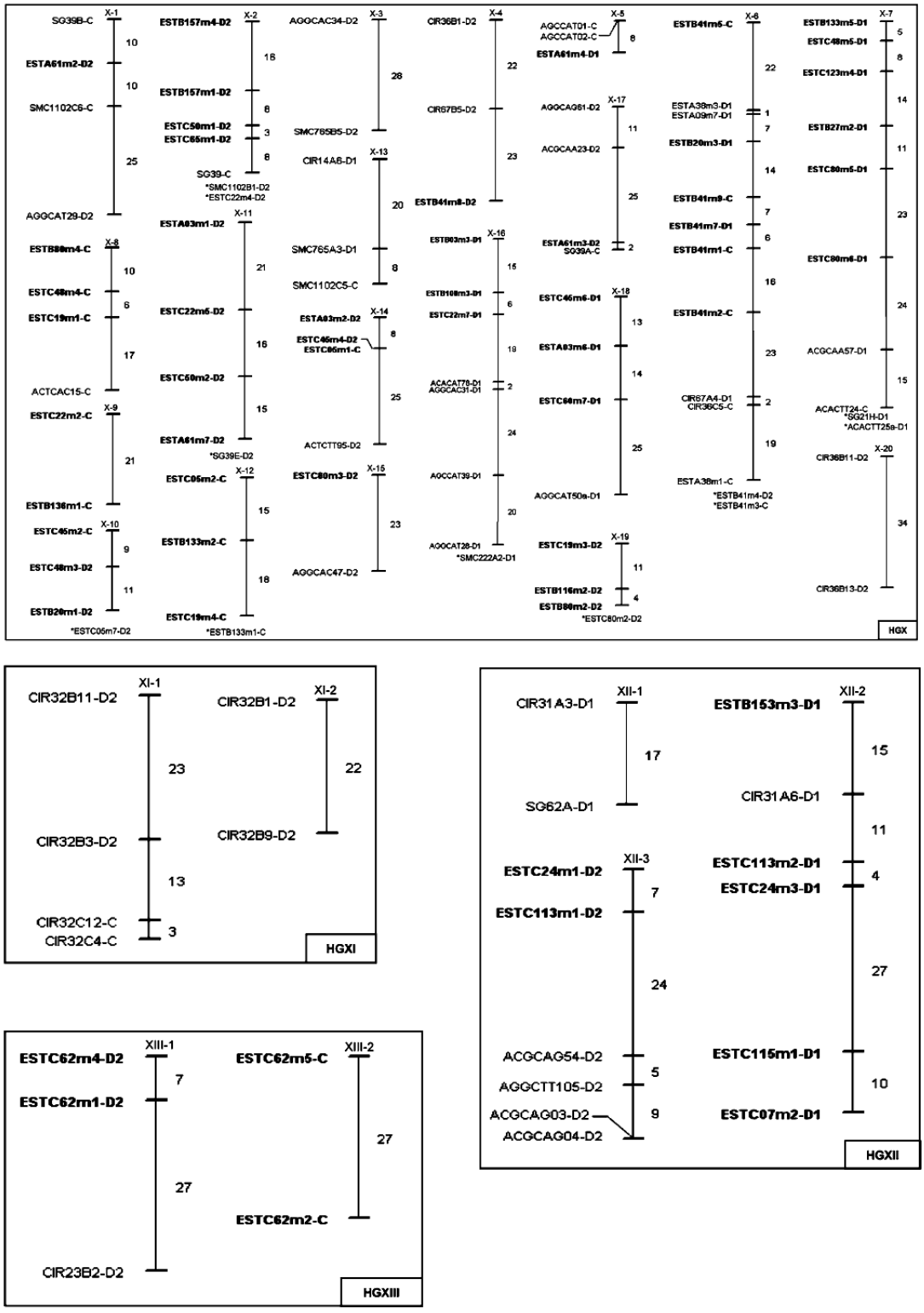


Fig. 1 continued

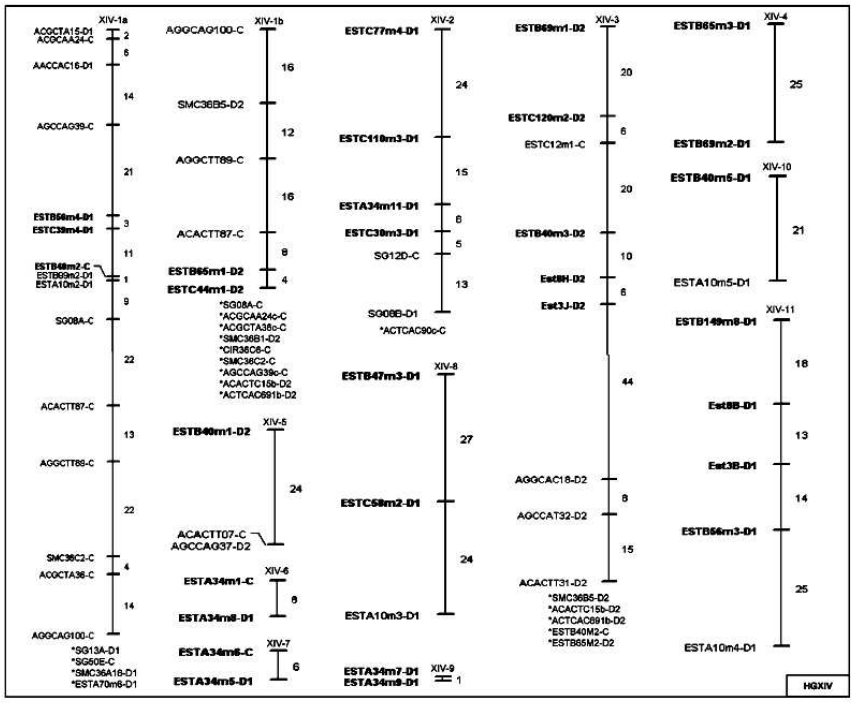


Fig. 1 continued

tent between CGs found within an HG. But, as large gaps were still often formed, more markers were needed for the precise positioning and ordering of these markers.

Homology groups in the present map were organized based on the groups of previous map. Three HGs (HGIV, HGVII and HGXI) were exactly the same on both this map and the previous one, and four HGs (HGV, HGXII, HGXIII and HGXIV) were established as new ones, based on EST-SSR markers. Except for HGIII and HGVI that were formed by two groups from the first map (HGIII and HGV, HGVI and HGXII, respectively), the others were basically the same with new co-segregation groups derived from EST-based markers or previous co-segregation groups with new linked markers (Fig. 1).

Twenty-four SSRs and three RFLPs detected duplicated loci within a co-segregation group. All but two homology groups (HGIV and HGXII) detected these duplications (Table 3). Apart from this, markers generated by seven EST-SSRs were mapped to different homology groups: ESTA09 (HGIII and HGX), ESTB03 (HGVI and X), ESTB20 (HGVIII and X), ESTB110 (III and VI), ESTB149 (HGI and HGII), ESTC57 (HGIII and HGVIII) and ESTC123 (HGV and HGX) (Table 3; Fig. 1).

Putative functions of the EST products containing SSRs and RFLPs

To exploite the potential utility of the EST-based markers, 113 EST-SSR loci and six EST-RFLP mapped probes were compared to genes of other species using BLASTX, directly from the SUCEST database. The EST was identified as the gene that showed the highest score among the candidate proteins. Blast result revealed that, amongst the EST-SSR sequences, 91.2% were derived from ESTs with functionally annotated hits in other organisms, predominantly *Arabidopsis thaliana*, followed by species taxonomically related to sugarcane such as, rice, maize, sorghum and other species. The remaining 8.8% showed no homology to sequences in the database interrogated (Table 3). From the 103 EST sequences

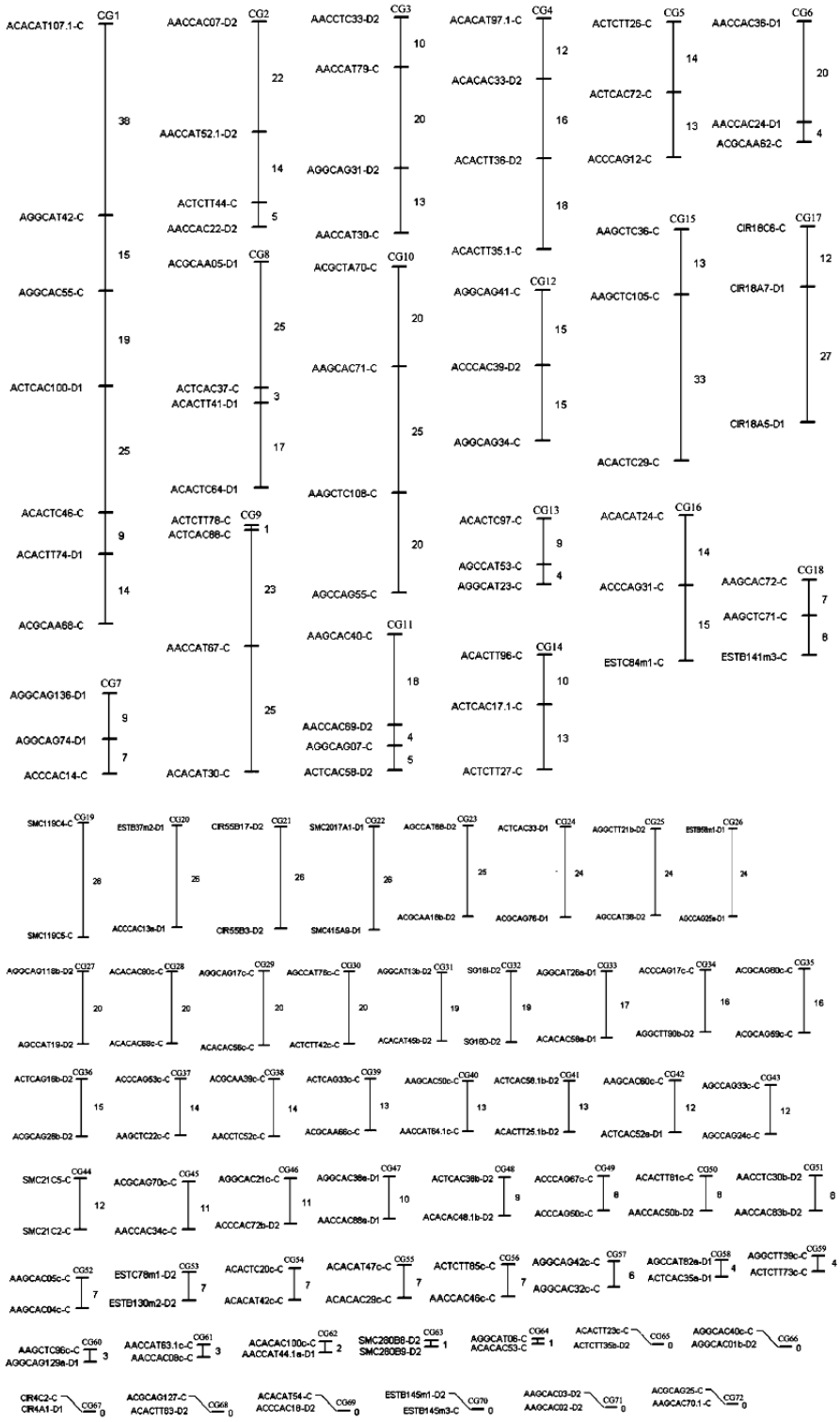


Fig. 1 continued

Table 3 EST-SSR and EST-RFLP linked markers: location on homology groups (HGs) and expected EST homology

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]	e-value
SCCCLR1001A05.g	III (2)	Est1	Est1CD1, Est1CC	*Sucrose synthase*	34391404	0.0
SCCCRZ1001B08.g	VI (4)	Est2	Est2BD1, Est2AD2, Est2DD2, Est2CD1	*Triosephosphate isomerase 1*	168647	e–125
SCCCLR1066B10.g	XIV (2)	Est3	Est3JD2, Est3BD1	*Cellulose synthase-4*	50935131	0.0
SCCCLR1001G06.g	VI (1)	Est7	Est7CD2	*S-adenosylmethionine synthetase*	51038250	0.0
SCCCLR1066B10.g	XIV (2)	Est8	Est8HD2, Est8BD1	*Cellulose synthase-4*	50935131	0.0
SCUTLR1058B04.g	VIII (1)	Est9	Est9A-D2	OSJNBa0072K14.5	50923903	0.0
SCCCCL7038A01.g	X (3)	ESTA03	m1D2, m2D2, m6D1	Putative *beta-mannan endohydrolase*	34909462	3.00E–39
SCEZLR1031G07.g	V (3)	ESTA06	m5C, m6D1, m4D1	*Lipid transfer protein*	25990344	2.00E–27
SCRUFL1118A11.b	III (2)	ESTA07	m2C, m1D2	Putative *apical-basal pattern formation protein*	50917563	1.00E–41
SCRLFL4027F03.g	III (3)	ESTA08	m3D1, m1D2, m2D2	Hypothetical protein	55773910	1.00E–16
SCCCRZ2C03C03.g	III / X (3)	**/*ESTA09	m3D2, m6D2, m5C, m7D1	No hit		
SCQGSB1143H06.g	XIV (4)	ESTA10	m2D1, m5D1, m3D1, m4D1	No hit		
SCRLAD1139G03.g	III (2)	ESTA14	m6C, m2D2, m3D2	No hit		
SCCCHR1004C07.g	III (5)	**ESTA15	m2C, m7D1, m3C, m4D2, m6D1, m5D1	*Caffeoyl CoA 3-O-methyltransferase*	32347451	2.00E–50
SCCCLR1001C05.g	III (1)	ESTA16	m2D1	Hypothetical protein	50900112	2.00E–97
SCEQRT2030G04.g	VI (3)	**ESTA17	m2D2, m1D2, m6D1, m3D2	Hypothetical protein	34894358	4.00E–54
SCVPCL6061A01.g	XIV (4)	** ESTA34	m11D1, m1C, m8D1, m6C, m5D1, m7D1, m9D1	Hypothetical protein	12597881	1.00E–66
SCJLRZ1025A05.g	X (1)	**ESTA38	m3D1, m1C	No hit		
SCJFLR1035E10.g	VI (1)	ESTA44	m1D1	OSJNBa0086B14.26	50924658	0.0
SCRULB2062C01.b	VI (3)	ESTA47	m1C, m3D2, m2D2	Putative *branched-chain amino acid aminotransferase protein*	34904468	6.00E–38
SCACAM2044B11.g	V (3)	ESTA48	m2D2, m1D2, m9D1	*Chloroplast phytoene synthase 1*	38037628	2.00E–76
SCEZHR1087D09.g	III (3)	ESTA49	m4C, m3D1, m2D2	Putative *glutathione S-transferase*	50918191	1.00E–68
SCSFHR1043F12.g	III (6)	ESTA53	m2C, m1D2,m5C, m8C, m4D2, m6C, m7D2	Indeterminate *spikelet 1*	2944040	1.00E–97
SCRULB1059B07.g	III (4)	**ESTA54	m4D2, m6C, m3D1, m2D2, m1D2	No hit		
SCQGST1032G11.g	VI (4)	ESTA55	m3D2, m1C, m4D2, m2D2	No hit		
SCSBSD1029G01.g	IX (3)	**ESTA58	m1C, m7D1, m6D1, m4D2, m3D2	No hit		
SCMCRT2103A12.g	X (4)	ESTA61	m2D2, m4D1, m3D2, m7D2	*Lipid transfer protein-like protein*	39748092	1.00E–24
SCCCLB1002D05.g	V (3)	ESTA63	m1C, m3D2, m2D2	*TPA: class III peroxidase 90 precursor*	55701047	e–146
SCMCRT2088D10.g	III (3)	ESTA66	m3C, m1D2, m5D1	Putative *protein kinase*	50901454	e–103
SCEQRT1024F02.g	I (2)	ESTA70	m6D1, m1C	*Peroxidase*	51038054	e–152
SCCCCL4013D09.g	VI / X (2)	*ESTB03	m4D1, m3D1	*12-oxo-phytodienoic acid reductase*	63021731	0.0
SCCCST1006B01.g	V (2)	ESTB07	m3D1, m1C	Putative *diphosphate-fructose-6-phosphate 1-phosphotransferase*	55168177	2.00E–56
SCSGLV1008C05.g	V (2)	ESTA68	m2C, m1C	*Fructose 1,6, bisphosphate aldolase*	40362980	3.00E–07
SCUTAM2009E01.g	IX (2)	**ESTB100	m5D1, m7D1, m1D2	*22 kDa drought-inducible protein*	49234816	2.00E–14
SCJFRZ2027D03.g	X (1)	ESTB108	m3D1	Hypothetical protein	50939565	4.00E–39

Table 3 continued

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]	e-value
SCCCCL4004A10.g	III / VI (2)	*ESTB110	m3D1, m2C	Putative *polyprotein*	48209910	4.00E–23
SCCCCL1001E03.g	V (2)	ESTB111	m2C, m4D1	No hit		
SCCCCL4004B01.g	X (1)	ESTB116	m2D2	P0497A05.15	50902306	8.00E–42
SCSBHR1053D01.g	III (2)	ESTB118	m3D1, m1D2, m2D2	*Homeobox transcription factor GNARLY1*	32351469	1.00E–16
SCCCST1002H09.g	I (3)	ESTB122	m9D2, m3D2, m6D2	Hypothetical protein	50931511	1.00E–32
SCMCAM1101H03.g	III (2)	ESTB131	m3D1, m2C	Putative *helicase*	50919223	2.00E07
SCRFAM1025G10.g	X (2)	ESTB133	m2C, m1C, m5D1	*HSV-1 stimulating-related protein-like*	57899937	6.00E–55
SCQGST3123C08.g	III (2)	ESTB134	m1C, m4D1	Putative *pectin methylesterase*	51090795	2.00E–29
SCEQRT2025G01.g	X (1)	ESTB136	m1C	Putative *AP2 domain containing protein*	50920393	1.00E–35
SCVPCL6044E08.g	III (1)	ESTB140	m2D1	Hypothetical protein	50910633	4.00E–22
SCCCCL4010B11.g	VIII (2)	ESTB146	m2D1, m1C	OSJNBa0071113.13	50929299	1.00E–51
SCSGFL4C06A02.g	XIV / II (2)	*ESTB149	m8D1, m6D1	*Cellulose synthase-7*	9622886	6.00E–54
SCUTLR1058B02.g	VI (1)	ESTB150	m3D1	OSJNBa0011P19.5	34894718	0.0
SCEZRZ3100H11.g	XII (1)	ESTB153	m3D1	*Multiple stress-associated zinc-finger protein*	37548823	2.00E–15
SCBGLR1119A10.g	VIII (1)	ESTB155	m1C	OSJNBa0043L09.23	50928999	2.0E–30
SCRFLR1034H10.g	X (1)	**ESTB157	m4D2, m1D2	Putative *thaumatin-like protein*	37531042	e–111
SCQGLR1041E11.g	V (1)	ESTB16	m2D2	SbCBF5	60593387	6.00E–62
SCQSRT2032H08.g	VIII / X (3)	*ESTB20	m2D1, m1D2, m3D1	Hypothetical protein	56784702	e–136
SCRLFL1006B02.g	V (1)	ESTB23	m1C, m3D2	MFS18	22647	8.00E–06
SCRUFL1120C04.b	VIII (1)	ESTB25	m2C	Putative *benzothiadiazole-induced somatic embryogenesis receptor kinase 1*	53792830	1.00E–13
SCRUSB1078F07.g	X (1)	ESTB27	m2D1	P0480C01.19	34910656	9.00E–43
SCVPLR2027E01.g	V (1)	ESTB35	m2C	MFS18	22647	3.00E–08
SCBGLR1002F11.g	III (4)	**ESTB39	m5D2, m4D2, m11D2, m3D2, m10D2,m9C, m2C, m8C, m1C	*Alpha 3 subunit of 20S proteasome*	34898416	e–127
SCBGLR1023A04.g	XIV (3)	ESTB40	m2C, m3D2, m5D1, m1D2	*Blue copper-binding protein-like*	50882442	3.00E–50
SCEQSD2077B12.g	X (2)	**ESTB41	m5C, m9C, m7D1, m1C, m2C, m4D2, m3C, m8D2	Hypothetical protein	50916060	5.00E–14
SCCCCL4007E05.g	V (2)	**ESTB43	m6D1, m2C, m1C	Putative *diphosphate-fructose-6-phosphate 1-phosphotransferase*	55168177	0.0
SCCCCL4013C08.g	VIII (1)	ESTB45	m4D1	OSJNBa0032F06.16	50929735	1.00E–58
SCCCLR1072G02.g	XIV (1)	ESTB47	m3D1	Putative *zinc finger transcription factor*	50933107	e–111
SCAGLR2011C01.g	XIV (2)	ESTB56	m4D1, m3D1	*Glycine-rich RNA-binding protein-like*	50934801	7.00E–83
SCJFRZ2013F12.g	IX (5)	**ESTB60	m8D2, m4D2, m5D2, m3C, m2C, m6C	OJ000114_01.16	50924506	1.00E–52
SCEZRZ1014F04.g	VI (4)	ESTB63	m3D2, m1C, m5D1, m3D1	*Mitogen activated protein kinase 6*	37594657	0.0
SCQSRT2033C08.g	IX (1)	ESTB64	m4D2	*RNA-directed DNA polymerase HMG-I and HMG-Y, DNA-binding*	87162631	e–117
SCCCLR1072G12.g	XIV (2)	ESTB65	m2D2, m1D2, m3D1	Putative *ribosomal protein L32*	50947773	9.00E–69
SCQGSD1045H12.g	IX (3)	**ESTB67	m8D2, m5D1, m2D1, m4D1, m1C	Putative *kaftirin preprotein*	22208459	7.00E–43
SCQGSD1045H12.g	IX (3)	ESTB68	m1D2, m2D2, m7D1	Putative *kaftirin preprotein*	22208459	7.00E–43
SCQGSD1045H12.g	XIV (2)	ESTB69	m1D2, m2D1	Putative *kaftirin preprotein*	22208459	7.00E–43

Table 3 continued

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]	e-value
SCJFRT1059E03.g	III (1)	ESTB75	m2C	Hypothetical protein	34902434	5.00E–25
SCJLLR1033A04.g	X (2)	ESTB80	m2D2, m4DC	*Ankyrin-like protein*	50902216	2.00E–24
SCSGFL4C02D07.g	VI (1)	ESTB82	m2C	Putative *peroxidase*	34913714	4.00E–53
SCEZRZ1014H07.g	IX (2)	ESTB92	m2D2, m1D2	Putative *sugar transporter*	22208506	0.0
SCCCFL3002B02.g	III (3)	ESTB94	m4C, m6D1, m2D2	No hit		
SCCCLR1C07C03.g	XIV (1)	ESTB99	m2D1	No hit		
SCEZHR1048C09.g	III (4)	ESTC01	m4D1, m5D1, m3C, m6D1	Putative *cysteine protease*	50918779	2.00E–52
SCUTST3087G12.g	X (3)	ESTC05	m1C, m2C, m7D2	Putative *cytochrome B5*	34905998	2.00E–67
SCEZFL4043H02.g	X (1)	ESTC07	m2D1	Hypothetical protein	9757905	1.00E–14
SCUTST3092H12.g	VI (4)	ESTC109	m4D1, m5D2, m1D2, m2D2	Hypothetical protein	34903888	5.00E–71
SCRFRZ3058E03.b	XIV (1)	ESTC110	m3D1	*ABA 8'-hydroxylase 1*	81362266	1.00E–20
SCCCST3C02F09.g	XII (1)	ESTC113	m2D1, m1D2	Putative *thaumatin-like protein*	50726592	2.00E–30
SCCCLR2004B05.g	I (1)	ESTC114	m1D2	*Phospholipase* putative	77548280	3.00E–44
SCBFRZ2017D04.g	XII (1)	ESTC115	m1D1	*U2AF small subunit*	68036691	6.00E–38
SCSBSB1095H03.g	V (1)	ESTC119	m1D1	Putative *leucine-rich repeat transmembrane protein kinase*	50905839	4.00E–15
SCCCLB1C04C09.g	XIV (1)	ESTC12	m1C	No hit		
SCEQLR1007E11.g	XIV (1)	ESTC120	m2D2	Putative *uracil phosphoribosyltransferase*	50906841	e-100
SCBGLR1027H03.g 5.00E–35	V / X (3)	**/	*ESTC123	m2C, m1C, m3D1, m4D1	mLIP15	1060935
SCSGLR1045E07.g	V (4)	ESTC129	m4C, m2C, m1D1, m5C	Putative *leucine-rich repeat transmembrane protein kinase*	50905839	3.00E–86
SCCCLR2C01B03.g	I (1)	ESTC13	m1D2	Putative *actin-depolymerizing factor*	34897072	5.00E–74
SCMCFL5004A10.g	III (1)	ESTC130	m2D2	*Glycine-rich RNA-binding protein*	10799202	7.00E–35
SCJFRZ2033G09.g	III (2)	ESTC131	m2D2, m1D2	*Glycine-rich RNA-binding protein*	10799202	1.00E–40
SCJLHR1027H06.g	III (1)	ESTC132	m1D2	*Glycine-rich RNA-binding protein*	10799202	7.00E–16
SCQSLB1049C05.g	III (1)	ESTC133	m1D2	*Glycine-rich protein*	2196542	5.00E–33
SCQGAM2026F07.g	III (1)	ESTC134	m2D2	*Glycine-rich RNA-binding protein*	10799202	4.00E–40
SCACLR1057A06.g	X (3)	ESTC19	m4C, m3D2, m1C	P0518C01.24	34906022	1.00E–10
SCBGLR1114E07.g	V (3)	ESTC21	m1D1, m2D1, m3C	*Lipid transfer protein*	311331	4.00E–34
SCJFRZ2033H11.g	X (4)	ESTC22	m4D2, m5D2, m2C, m7D1	Hypothetical protein	50939961	8.00E–54
SCQGST1032E05.g	XI (2)	ESTC24	m3D1, m1D2	*General transcription factor TFIIB*	18481632	e-175
SCJFRT2060G09.g	VIII (3)	ESTC25	m2D1, m1C, m3D1	CAA303719.1 protein	5777631	0.0
SCCCRZ2C01E09.g	XIV (1)	ESTC30	m3D1	*Mitochondrial uncoupling protein 4*	51860691	e-143
SCJLRT3077F09.b	XIV (1)	ESTC39	m4D1	Putative *glycosyltransferase*	50939609	4.00E–26
SCCCRZ1C01B10.g	V (2)	ESTC42	m1D2, m2D2, m3D1	Putative *receptor-like kinase Xa21-binding protein 3*	52353644	e-127
SCQGST1034A05.g	XIV (1)	ESTC44	m1D2	B1066G12.15	34904540	3.00E–83
SCMCST1052A09.g	X (3)	ESTC45	m4D2, m6D1, m2C	SbPCL1	71067066	2.00E–47
SCCCCCL3005C08.b	III (2)	ESTC47	m3D1, m1D2	Putative *acetyl-CoA synthetase*	26451642	e-155
SCEPFL4177D11.g	X (3)	ESTC48	m3D2, m4C, m5D1	P0518C01.24	34906022	3.00E–12

Table 3 continued

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]	e-value
SCEPLR1051G11.g	V (1)	ESTC49	m1C	mLIP15	1060935	2.00E–44
SCEZRZ3050D12.g	X (2)	ESTC50	m1D2, m2D2	At1g68060/T23K23_9	23463063	3.00E–29
SCACRZ3108F03.g	III / VIII (4)	**/*ESTC57	m3D2, m8C, m1D2, m6D1, m7D1	OSJNBa0072K14.5	50923903	2.00E–26
SCVPLR1049C03.g	XIV (1)	ESTC58	m2D1	*Methylmalonate semi-aldehyde dehydrogenase*	50934827	0.0
SCCCLR1C04G09.g	X (1)	ESTC60	m7D1	*NADP-specific isocitrate dehydrogenase*	34911932	0.0
SCMCFL5009H03.g	XIII (1)	**ESTC62	m4D2, m1D2, m5C, m2C	*Glyceraldehyde-3-phosphate dehydrogenase*	474408	1.00E–38
SCJLFL3018F04.g	X (1)	ESTC65	m1D2	*Galactosyl transferase GMA12/MNN10 family* putative	77551427	2.00E–26
SCRURT2009A09.g	XIV (1)	ESTC77	m4D1	Putative *caffeoyl-CoA O-methyltransferase 1*	50947279	4.00E–28
SCJFST1016B06.g	X (3)	**ESTC80	m2D2, m5D1, m6D1, m3D2	*Glycosyltransferase*	63087728	8.00E–82
SCJLLR1101C01.g	III (4)	ESTC81	m3C, m2C, m1D2, m4D2	Similar to *ubiquinol cytochrome c reductase*	34898476	1.00E–33
SCAGLR1021D10.g	I (2)	ESTC91	m1D2, m2D2	*Putative fructose-1,6-bisphosphatase*	50931573	1.00E–15

[a] EST *clusters* of SUCEST database (http://sucest.lad.ic.unicamp.br/en/)

[b] Homology groups (HGs)—indicated in *parentheses* the number of co-segregation groups (CGs) where each EST-derived markers was assembled

[c] EST-RFLPs: Est1, Est2, Est3, Est7, Est8 and Est9. EST-SSRs: ESTA, ESTB and ESTC refer, respectively, to di-, tri- and tetranucleotide primer pairs

[d] For EST-SSRs: mark names present allele number and their type of cross (D1, D2 or C), while for the EST-RFLPs, mark names are the same as described in the *materials and methods* section

[e] Best hit description

[f] GenBank identification

* EST-SSR duplicated between HGs

** EST-SSR duplicated in CGs. Other type markers also duplicated in CGs: SG50(HGI), SG94 and SG41(HGII), CIR35(HGVI), CIR21(HGVII), CIR36(HGX), CIR32(HGXI), SMC236(HGVIII) and SMC1047(HGIX)

homologous to functional hits, only 13 were associated with hypothetical or unknown protein. In the majority of the EST-SSRs, the BLAST homology comparisons indicated that the sequences had characterized functions.

Some important enzymes related to sugarcane metabolism were identified. They included putative *sugarcane transporter* (*cluster* SCEZRZ1014H07.g), *fructose-1, 6-bisphosphate adolase* (*cluster* SCSGLV 1008C05.g), *peroxidase* (*clusters* SCCCLB1002D05.g, SCEQRT1024F02.g, SCSGFL4C02D07.g), *fructose–1,6-bisphosphatase* (*clusters* SCAGLR 1021D10.g, SCSGLV1008C05.g), *diphosphate-fructose-6-phosphate 1-phosphotransferase* (*clusters* SCCCCL4007E05.g, SCCCST1006B01.g) and *glyceraldehyde-3-phosphate dehydrogenase* (*cluster* SCMCFL5009H03.g). Besides the EST-SSRs, the EST-RFLP probes also directly mapped genes such as: *sucrose synthase* (*cluster* SCCCLR1001A05.g), *triosephosphate isomerase 1* (*cluster* SCCCRZ 1001B08.g), *cellulose synthase-4* (*cluster* SCCCLR 066B10.g) and *S-adenosylmethionine synthetase* (*cluster* SCCCLR1001G06.g).

Discussion

The authors have exploited the SUCEST database to generate a functionally associated marker-based genetic linkage map for a sugarcane mapping population derived from a cross between two pre-commercial cultivars (SP80-180 and SP80-4966). Expressed sequences from sugarcane that had been classified based on sequence analysis were used to develop EST-RFLP and EST-SSR markers. These EST-based markers were located in the previous SP80-180 × SP80-4966 population genetic map (Garcia et al. 2006). The coverage and marker density of this genetic map was increased with the integration of these EST-derived markers. The map, generated using the method proposed by Wu et al. (2002), has 307 more markers than the previous one and enhanced genome coverage in 3.658,7 cM with the new EST-based CGs. The total length of the current map is 6.261,1 cM with a substantial percentage of unlinked markers (60%). The map is clearly not saturated, given the small size of many CGs and the numerous unlinked markers. Besides, the number of CGs (192 CGs) exceeds chromosome number expected for modern cultivars ($2n$ = 100–130). This indicates that gaps remain on most if not all chromosomes, showing that the map is already incomplete. Since there is a constraint to discard markers in multiples doses, i.e., duplex of monoparental origin, triplex or higher multiplex markers, these gaps are evidently expected.

The number of unlinked markers is smaller than the number obtained in the previous genetic map of this population (68%), however it is higher than that obtained in other maps for sugarcane (Reffay et al. 2005; Aitken et al. 2005; Raboin et al. 2006). Nevertheless, the current map was generated from a cross between two interspecific modern sugarcane cultivars, aneuploids and with a very complex genetic system, what makes difficult linkage between markers and, as a result, the loss of informative markers for mapping can be expected. The genome of modern cultivar needs to be understood in relation to the genome composition of its *S. officinarum* ($2n$ = 80) and *S. spontaneum* ($2n$ = 40–128) progenitors. Structural differences are present between the two genomes as differences in base number are observed (Butterfield et al. 2001). Actually, chromosome-pairing behavior has not been definitely clarified and information on the structure and organization of the genome has been largely speculative. Thus, difficulties in mapping are expected in this type of population.

Comparison with other sugarcane results (Reffay et al. 2005; Aitken et al 2005; Raboin et al. 2006) has shown that a higher number of double single-dose markers (52% of segregating markers) segregated in the SP80-180 × SP80-4966 population. Indeed, SP80-180 and SP8049-66 have some ancestors in common in the first generations, for instance, POJ2364 [Kassoer × POJ100]. It is known that today's sugarcane cultivars are almost exclusively backcross derivatives involving a very small number of parental clones (Price 1965). Therefore, a narrow genetic base of modern hybrid varieties is surely one of the principal causes of the present slow rate of sugarcane breeding progress (Berding and

Roach 1987). Thus, taking this into consideration, it is to be expected that populations generated from crosses between modern cultivars present a high number of double single-dose markers.

Although less informative, the C type marker can establish linkage relationship between markers originated from each parent (D1 and D2 types). In fact, to construct an integrated genetic map in sugarcane, a large number of these markers should be available (Garcia et al. 2006). Hence, the high number of C type markers in our population enhanced the construction of the integrated map. However, as in the previous map, linkage between groups with D1 markers (SP80-180 origin) and with D2 markers (SP80-4966 origin), through C markers, was observed in only a few groups: only in CGVIII-2, D1 type markers were linked with D2 type markers; the co-segregation groups CGXIV-1, CGIII-3, CGIII-4, CGVI-1 were also made up of all marker types, however, due to insufficient linkage information, they were shared in two groups ('a' and 'b'). Due to the same reason, CGIX-1 was split into two groups (CGIX-1a and CGIX-1b), albeit only D2 and C type markers were assembled.

Previous maps of sugarcane cultivars have identified ten (Grivet et al. 1996), eight (Reffay et al. 2005; Aitken et al 2005) and seven (Rossi et al. 2003) homology groups. But for all of them, homology groups were established separately from each parent. In this study, co-segregation groups of both parents were used to assign the fourteen putative homology groups of this integrated map. Differences in chromosome structure between the progenitor species and pairing behavior in modern varieties suggest that in sugarcane hybrids, the hybrid monoploid number is likely to be greater than 10 (Butterfield et al. 2001). As HGs represent a basic number of chromosomes, the number of HGs achieved in this study is in agreement with the estimated number. The 14 HGs showed differences in coverage and in the CG number across them. However, due to aneuploidy, an unequal number of chromosomes in each HG is likely to occur. The discrepancy of coverage between HGs highlights the difficulty in mapping large parts of the genome.

Some CGs contained duplicated loci. Little is known about duplication of genome segments within the monoploid chromosomes of *Saccharum*, since mapping strategy on single-dose markers hinders the identification of the duplicated regions. However, duplicated regions of the genome could be identified within CGs and between HGs, in the current map. These regions remain undetected in relatively low-density maps, but in many sugarcane maps, they were mentioned. Ming et al. (1998) detected duplications within and across the co-segregation groups. Jannoo et al. (1999) observed eight cases of duplication to different linkage groups in R570. The proportion of duplicated regions in the monoploid hybrid genome is potentially higher than in ancestral species.

None of the published genetic maps of sugarcane are saturated. No more than one third of the sugarcane genome is estimated to be tagged on, even in the most refined maps, due to the genetic complexity of its genome. To date, the sugarcane maps were constructed for different population types, with anonymous markers. Due to the fact that genetic maps based on expressed sequences have been constructed on a variety of other agronomically important species such as rice (Wang et al. 2005), maize (Falque et al 2005) and cotton (Han et al. 2006), these markers may be used to construct functional sugarcane maps.

The value of EST-SSRs is enhanced by their superior transferability across taxon boundaries, as demonstrated in grape (Decroocq et al. 2003) and in sugarcane (Cordeiro et al. 2001) as is their potential as 'perfect markers' for functionally defined genes involved in determining agronomic traits. In terms of the discovery of SSR, development of functional primer pairs, and polymorphism detection, EST-SSR markers are less efficient than DNA-derived SSRs, however these caveats are balanced by the relatively low expense with EST-SSR development as a product of a genomic database (Faville et al. 2004).

The fact that ESTs containing SSRs exhibited sequence similarity to genes with a wide range of functions suggests that there is potential to identify EST-SSRs that may be directly involved in determining agronomically important traits. For example, there are 15 ESTs with strong homology to sugarcane metabolism proteins. The identification of genes that determine economically important

plant traits provides important tools to further manipulate plant function and performance, through enhanced conventional breeding using the gene's DNA sequence as a 'perfect' marker for trait selection. The development of a functionally defined gene-based genetic map of sugarcane provides the basis for the correlation of molecular variation associated with functional sequences with the locations of QTLs for putatively related traits. QTL analysis using EST-RFLP and EST-SSR allows the identification of associations between functionally associated marker locations and QTLs. However, a large number of markers are necessary to build a genetic map and to obtain sufficient genome coverage for QTL analysis, due to the large number of chromosomes in sugarcane. Although the current map is not well saturated, the genome coverage of this map will probably facilitate the detection of QTLs for important traits in sugarcane, because the addition of known-function gene markers could greatly enhance the utility of a genetic map, as it facilitates the transition from genetic linkage analysis to a candidate gene mapping approach so as to dissect complex traits.

Acknowledgements This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 02/01167-1) and from CTC (Centro de Tecnologia Canavieira – Piracicaba/SP). K.M.O. and M.M.P. received Doctorate fellowships, respectively, from FAPESP (02/00197-4) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); K.M.O. also received PDEE fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, 0181-05-6); L.H.M. received an MS fellowship from FAPESP (03/07960-6); TGM and GRAM received IC fellowships from FAPESP (04/10596-9 and 05/59268-6, respectively); L.R.P. received a Postdoctorate fellowship from FAPESP (01/14656); A.P.S., A.A.F.G. and A.F. received a research fellowship from CNPq. The authors would like to thank Dr. Angélique D'Hont (*Centre de Coopération Internationale en Recherche Agronomique pour le Développement*) for valuable discussions and anonymous reviewers for carefully reading the manuscript.

References

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110:789–801

Berding N, Roach BT (1987) Germplasm collection, maintenance and use. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier press, Amsterdam, pp 143–210

Butterfield M, D'Hont A, Berding N (2001) The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc Soc Afr Sugarcane Technol Assoc 75:1–5

Carson DL, Botha FC (2000) Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769–1779

Casu R, Dimmock C, Thomas M, Bower N, Knight D, Grof C, McIntyre L, Jackson P, Jordan D, Whan V et al (2001) Genetic and expression profiling in sugarcane. Proc Int Soc Sugarcane Technol 24:626–627

Cato SA, Gardner RC, Kent J, Richardson TE (2001) A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102:296–306

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248–1257

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp) ESTs cross transferable to *Erianthus* and *Sorghum*. Plant Sci 160:1115–1123

Creste S, Tulmann-Neto A, Figueira A (2001) Detection of single repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. Plant Mol Biol Rep 19:229–306

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterization of the double-genome structure of modern sugarcane cultivars (*Saccharum* ssp) by molecular cytogenetics. Mol Gen Genet 250:405–413

Da Silva JAG, Honeycutt RJ, Burnquist WL, Al-Janabi SM, Sorrells ME, Tanksley S, Sobral B (1995) *Saccharum spontaneum* L. 'SES 208' genetic linkage map containing RFLP and PCR based markers. Mol Breed 1:165–179

Decroocq V, Favé M, Hagen L, Bordenhave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106:912–922

Doerge RW (1996) Constructing genetic maps by rapid chain delineation. J Agric Genom 2:6

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene Loci. Genetics 170(4):1957–1966

Faville MJ, Vecchies AC, Schreiber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT, Forster JW (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110:12–32

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, da Silva JAG, Ulian EC, Figueira A, Souza A (2006) Development of an integrated genetic map of a sugarcane (*Saccharum*

spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet 112:298–314

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp): genome organization in a highly polyploid and anueploid interspecific hybrid. Genetics 142:987–1000

Guimarães CT, Honeycutt RJ, Sills GR, Sobral BWS (1999) Genetic maps of *S. officinarum* L and *S. robustum* Brandes & Jew. Ex Grassl. Genet Mol Biol 22:125–132

Han ZG, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Grossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112:430–439

Hoarau J-Y, Souza Glaucia, D'Hont A, Menossi M, Pinto LR, Souza AP, Grivet L, Menck CFM, Ulian EC, Vincentz M (2006) Sugarcane, a tropical crop with a highly complex genome. In: Morot-Gaudry JF, Lea P, Briat JF (Org.). Plant genomics (in press)

Hoisington D, Khairallah M, González-de-Léon D (1994) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. CYMMYT, México

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among sugarcane cultivars. Theor Appl Genet 99:1053–1060

Kantley RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eug 12:172–175

Ming R, Liu SC, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore P, Burnquist W, Sorrells ME, Irvine J, Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Pinto LR, Oliveira KM, Marconi TG, Garcia AAF, Ulian EC, Souza AP (2006) Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed 125:378–384

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, Souza AP (2004) Survey in the expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47:795–804

Price S (1965) Interespecific hybridizations in sugarcane breeding. Proc Int Soc Sugarcane Technol 12:1021–1026

Raboin L-M, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau J-Y, D'Hont A (2006) Genetic mapping in the high polyploid sugarcane using a bi-parental progeny: identification of a gene controlling stalk colour and new rust resistance gene. Theor Appl Genet 112(7):1382–1391

Reffay N, Jackson PA, Aitken KS, Hoarau J-Y, D'Hont A, Besse P, McIntyre CL (2005) Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. Mol Breed 15:367–381

Roach BT (1969) Cytological studies in *Saccharum*. Chromosome transmission in interspecific and intergeneric crosses. Proc Int Soc Sugar Cane Technol 13:901–920

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys M-A, D'Hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Gen Genomics 269:406–419

Vettore AL, Da Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Gen Mol Biol 24:1–7

Vettore AL, da Silva, FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Giglioti EA, Lemos MV, Coutinho LL, Nobrega MP, Carrer H, Franca SC, Bacci-Junior M, Goldman MH, Gomes SL, Nunes LR, Camargo LE, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon ML, Ferro JA, Silveira HC, Marini DC, Lemos EG, Monteiro-Vitorello CB, Tambor JH, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MM, de Rosa Junior VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SM, Nobrega FG, Menck CF, Braga MD, Telles GP, Cara FA, Pedrosa G, Meidanis J, Arruda P (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725–2735

Wang XS, Zhu J, Mansueto L, Bruskiewich R (2005) Identification of candidate genes for drought stress tolerance in rice by the integration of a genetic (QTL) map with the rice genome physical map. J Zhejiang Univ Sci B 6(5):382–388

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl Genet 83:294–300

Wu R, Ma CX (2002) Simultaneous maximum likelihood estimation of linkage and linkage phase in outcrossing species. Theor Pop Biol 61:349–363

ARTIGO III

"Characterization of new polymorphic functional markers for sugarcane"

K.M Oliveira, L.R. Pinto, T.G. Marconi, M. Mollinari, E.C. Ulian, S.M. Chabregas, M.C. Falco, W. Burnquist, A.A.F. Garcia, A.P. Souza

Publicado na revista

Genome (52: 191-209; 2009)

Characterization of new polymorphic functional markers for sugarcane

K.M. Oliveira, L.R. Pinto, T.G. Marconi, M. Mollinari, E.C. Ulian, S.M. Chabregas, M.C. Falco, W. Burnquist, A.A.F. Garcia, and A.P. Souza

Abstract: Expressed sequence tags (ESTs) offer the opportunity to exploit single, low-copy, conserved sequence motifs for the development of simple sequence repeats (SSRs). The authors have examined the Sugarcane Expressed Sequence Tag database for the presence of SSRs. To test the utility of EST-derived SSR markers, a total of 342 EST–SSRs, which represent a subset of over 2005 SSR-containing sequences that were located in the sugarcane EST database, could be designed from the nonredundant SSR-positive ESTs for possible use as potential genic markers. These EST–SSR markers were used to screen 18 sugarcane (*Saccharum* spp.) varieties. A high proportion (65.5%) of the above EST–SSRs, which gave amplified fragments of foreseen size, detected polymorphism. The number of alleles ranged from 2 to 24 with an average of 7.55 alleles per locus, while polymorphism information content values ranged from 0.16 to 0.94, with an average of 0.73. The ability of each set of EST–SSR markers to discriminate between varieties was generally higher than the polymorphism information content analysis. When tested for functionality, 82.1% of these 224 EST–SSRs were found to be functional, showing homology to known genes. As the EST–SSRs are within the expressed portion of the genome, they are likely to be associated to a particular gene of interest, improving their utility for genetic mapping; identification of quantitative trait loci, and comparative genomics studies of sugarcane. The development of new EST–SSR markers will have important implications for the genetic analysis and exploitation of the genetic resources of sugarcane and related species and will provide a more direct estimate of functional diversity.

Key words: sugarcane, simple sequence repeat (SSR), expressed sequence tag (EST), functional markers.

Résumé : Les étiquettes de séquences exprimées (EST) offrent une opportunité d'exploiter des motifs conservés présents en simple ou peu de copies pour le développement de microsatellites (SSR). Les auteurs ont inspecté la base de données d'EST de la canne à sucre pour la présence de SSR. Pour vérifier l'utilité des SSR dérivés d'EST, 342 EST–SSR, lesquels représentent un sous-ensemble des 2 005 séquences contenant des SSR qui ont été trouvées au sein de cette banque d'EST de la canne à sucre, étaient propices au développement de marqueurs génétiques parmi les EST non-redondants contenant des EST. Ces marqueurs EST–SSR ont été employés pour cribler 18 variétés de canne à sucre (*Saccharum* spp.). Une proportion élevée (65,5%) de ces EST–SSR qui ont amplifié des produits de la taille attendue étaient polymorphes. Le nombre d'allèles variait entre 2 et 24, pour une moyenne de 7,55 allèles par locus, et les valeurs de contenu d'information de polymorphisme variaient entre 0,16 et 0,94, pour une moyenne de 0,73. La capacité de tels jeux de marqueurs à discriminer entre les variétés était généralement plus élevée que l'analyse de contenu d'information de polymorphisme. Lorsque testés pour leur fonction, 82,1% de ces 224 marqueurs EST–SSR se sont avérés fonctionnels en montrant de l'homologie avec des gènes connus. Puisque ces EST–SSR font partie de la portion exprimée du génome, ils sont vraisemblablement associés à des gènes d'intérêt, ce qui augmente leur utilité pour des fins de cartographie génétique, pour l'identification de locus de caractères quantitatifs et pour des études de génomique comparative chez la canne à sucre. Le développement de nouveaux marqueurs EST–SSR aura d'importantes retombées en matière d'analyse génétique et d'exploitation des ressources génétiques chez la canne à sucre et des espèces apparentées en plus de fournir un estimé plus direct de la diversité fonctionnelle.

Received 13 March 2008. Accepted 21 October 2008. Published on the NRC Research Press Web site at genome.nrc.ca on 23 January 2009.

Corresponding Editor: B. Golding.

K.M. Oliveira, T.G. Marconi, and A.P Souza.[1] Centro de Biologia Molecular e Engenharia Genética (CBMEG) – Universidade Estadual de Campinas (UNICAMP), Departamento de Genética e Evolução, Instituto de Biologia, Cidade Universitária Zeferino Vaz, CP 6010, CEP 13083-875, Campinas-SP, Brasil.
L.R. Pinto. Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana – IAC/Apta, Anel Viário Contorno Sul, Km 321, CP 206, CEP 14.001-970, Ribeirão Preto-SP, Brasil.
M. Mollinari and A.A.F. Garcia. Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba-SP, Brasil.
E.C. Ulian, S.M. Chabregas, M.C. Falco, and W. Burnquist. Centro de Tecnologia Canavieira – CTC, Caixa Postal 162, 13400-970, Piracicaba, São Paulo, Brasil.

[1]Corresponding author (e-mail: anete@unicamp.br).

Mots-clés : canne à sucre, microsatellite (SSR), étiquette de séquence exprimée (EST), marqueurs fonctionnels.

[Traduit par la Rédaction]

Introduction

Among different classes of molecular markers, simple sequence repeat (SSR) markers are useful for a variety of applications in plant genetics and breeding because of their reproducibility, multialelic nature, codominant inheritance, relative abundance, and good genome coverage (Powell et al. 1996*a*). Indeed, in recent years, they have become the most chosen marker for population genetic analyses (Pashley et al. 2006). In general, SSRs are identified from either genomic DNA or cDNA sequences. However, the conventional isolation and characterization of such markers via genomic DNA libraries are costly and time consuming. To reduce cost and time, various clone procedures to create libraries enriched for SSRs have been developed (Edwards et al. 1996). Nevertheless, genomic SSR marker development is still laborious (Gupta and Varshney 2000). Therefore, over the past decade, there has been an exponential increase in the availability of DNA sequence data from a wide variety of taxa, including a wealth of expressed sequence tags (ESTs) that are typically single-pass sequences produced from cDNAs and are one way of overcoming the problem of cost to become a promising alternative to the development of traditional "anonymous" SSRs following standard methods.

Moreover, SSR markers derived from EST sequences have some intrinsic advantages over genomic SSR because they are quickly obtained by electronic sorting and are present in expressed regions of the genome. The usefulness of these genic SSRs also lies in their expected transferability because the primers are designed from the more conserved coding regions of the genome (Varshney et al. 2005*a*). These EST–SSRs are useful as anchor markers for comparative mapping and evolutionary studies and for assaying the functional diversity in natural populations and germplasm collections. Furthermore, as they represent transcribed genes, a putative function can be deduced by a homology search (Varshney et al. 2005*a*). Thus, the relative cheapness of EST sequencing and its associated automation often make EST sequencing the most attractive route for broad sampling of the transcriptome (Rudd 2003).

This alternative strategy to develop SSR markers is based on the search for EST-derived SSRs from cDNAs deposited in the public databases using data mining pipelines composed primarily of SSR search and primer design programs. With recent increasing emphasis on functional genomics, large data sets of ESTs are being developed, and with evolving bioinformatic tools, it is now possible to identify and develop EST–SSR markers on a large scale in a time- and cost-effective manner. For example, ESTs have been scanned for the presence of SSRs in cotton (Han et al. 2004), barley (Thiel et al. 2003), maize (Kantety et al. 2002; Sharopova et al. 2002; Varshney et al. 2002), rice (Kantety et al. 2002), rye (Khlestkina et al. 2004), sorghum (Kantety et al. 2002; Varshney et al. 2002), wheat (Gupta et al. 2003; Nicot et al. 2004), sunflower (Pashley et al. 2006), citrus (Chen et al. 2006), coffee (Bhat et al. 2005; Aggarwal et al. 2007), and even in sugarcane (Cordeiro et al. 2001; Pinto et al. 2004). Thus, the generation of EST-derived SSR markers has become an attractive complement to existing SSR collections.

Some studies suggested that EST–SSR markers could be used in related plant species for which little information on SSRs or ESTs (Cordeiro et al. 2001; Rossetto et al. 2002; Scott et al. 2000; Thiel et al. 2003; Bhat et al. 2005; Varshney et al. 2005*b*) is available. These markers are also good candidates for the development of conserved orthologous markers for genetic analysis and breeding of different species (Varshney et al. 2005*b*).

EST-derived cluster sequences have been widely annotated with tentative functions. The annotations are homology based and EST sequences or clusters inherit the annotative attributes of their match. Although individual EST collections might only represent as much as 60% of the host genes, the overlap between collections provides a suitable hunting ground for the selection of the common genes (Rudd 2003). The characterization of genetic variation within natural populations and among breeding lines is crucial for effective conservation and exploitation of genetic resources for crop improvement programs. Evaluation of germplasm with EST-derived markers might enhance the role of genetic markers by assaying the variation in transcribed and known-function genes, although there is a higher probability of bias owing to selection (Varshney et al. 2005*a*).

In the absence of complete genome sequences, the desire to generate high-density maps of the different genomes remains a priority for the directed identification of specific genes. In recent years, the EST–SSR loci have been integrated, or genome-wide genetic maps have been prepared, in several plant species such as barley (Pillen et al. 2000; Thiel et al. 2003), cotton (Han et al. 2006), kiwifruit (Fraser et al. 2004), raspberry (Graham et al. 2004), rice (Temnykh et al. 2000; Khlestkina et al. 2004), ryegrass (Warnke et al. 2004), tall fescue (Saha et al. 2004), wheat (Gao et al. 2004; Nicot et al. 2004; Yu et al. 2004), white clover (Barrett et al. 2004), and sugarcane (Raboin et al. 2006; Oliveira et al. 2007).

Sugarcane (*Saccharum* spp.) is a genetically complex polyploid grass, which makes the identification of associations between genes and traits difficult. For crop plants with complex genomes such as sugarcane, gene characterization is currently best achieved via EST analysis where sequence information is restricted to genes that are actually functioning in a particular tissue or situation (Casu et al. 2005). In the past few years, several projects for the sequencing of sugarcane ESTs have been initiated in South Africa (Carson and Botha 2000), Brazil (http://sucest.lad.ic.unicamp.br/en/), and Australia (Casu et al. 2001), allowing the development of EST–SSRs for this species (Cordeiro et al. 2001; Pinto et al. 2004).

Published by NRC Research Press

Table 1. Identification of the 18 sugarcane genotypes used for marker validation with their pedigree relationships covering two generations and their origin.

Genotype	Pedigree	Origin
CB36-24	(POJ2364 × EK28) POJ2878 × ?	Campos, Brazil
CB40-77	(POJ2364 × EK28) POJ2878 × CO290 (CO221 × D74)	Campos, Brazil
SP79-1011	(CO419 × CO419) NA5679 × CO775 (POJ2878 × CO371)	São Paulo, Brazil
SP70-1284	(POJ2878 × ?) B4176 × ?	São Paulo, Brazil
SP71-1406	(CO419 ×× CO419) NA5679 × ?	São Paulo, Brazil
SP80-3280	(CP5530 × CP5376) SP71-1088 × H575028 (H49134 × ?)	São Paulo, Brazil
SP79-6134	(H53263 × H507209) H634644 × ?	São Paulo, Brazil
SP79-2312	(SP71-6106 × ?) CP5659 × ?	São Paulo, Brazil
SP80-4966	(NA5679 × ?) SP711406 × ?	São Paulo, Brazil
SP80-180	(B3337 × ?) B3337 × ?	São Paulo, Brazil
IAC51-205	(POJ2364 × EK28) POJ2878 × ?	Campinas, Brazil
IAC64-257	(POJ2878 × CO290) CO419 × IAC49-131 (CP27108 × ?)	Campinas, Brazil
RB739359	(CO419 × MZ336) IANE5534 × ?	Republic of Brazil
RB855035	(CP521 × CP48103) L6014 × SP701284 (CB4176 × ?)	Republic of Brazil
RB855536	(IAC48–65 × ?) SP70-1143 × RB72454 (CP5376 × ?)	Republic of Brazil
Gandacheni (*S. barberi*)	*S. barberi* × ?	Saretha, India
IJ76-314 (*S. officinarum*)	*S. officinarum* × ?	Iryan, Java
Maneria (*S. sinense*)	*S. sinense* × ?	Pansahi, China

Our group has been using the Brazilian EST database (Sugar Cane EST Project (SUCEST)) to exploit that available source for EST-derived marker development. As a first effort, Pinto et al. (2004) carried out a survey for SSRs in the SUCEST database and characterized 30 EST–SSRs, generating an initial set of markers to be used in diversity and mapping analysis, as part of an attempt to generate molecular genetic markers for sugarcane. In a second work, Pinto et al. (2006) reported the usefulness of those markers in diversity analysis, comparing the EST–SSRs with SSRs derived from genomic libraries (gSSRs). The authors revealed the polymorphism levels of the EST–SSR markers, verified their discrimination power for fingerprinting, and compared the polymorphism between them and gSSR markers considering a common subset of sugarcane clones. The results demonstrated that EST–SSRs could be successfully used for genetic relationship analysis, extending the knowledge of sugarcane genetic diversity to a functional level. More recently, Oliveira et al. (2007) constructed a sugarcane genetic linkage map that is populated by 149 functionally associated markers. All of the work developed by our group has revealed a total of 372 EST–SSRs for sugarcane. However, from those EST–SSRs, only 30 have been characterized (Pinto et al. 2004). Thus, in the present article, we describe the characterization of the 149 mapped EST–SSRs (Oliveira et al. 2007) and of the other 193 EST–SSR markers developed and never before presented. Thus, the objectives of this paper are to (*i*) show the results of the usage of the Brazilian EST database, developed as part of SUCEST, to characterize SSR markers for sugarcane, stating their primer amplification conditions, polymorphism information content (PIC), discrimination power (DP), and putative functions of their predicted products, (*ii*) report 224 EST–SSR markers evaluating them for polymorphism and genetic diversity analysis across 18 sugarcane varieties and also their usefulness for sugarcane genotyping, arriving at a set of robust microsatellite markers for this crop, and (*iii*) compare the results with previous analysis using AFLP and pedigrees (Lima et al. 2002).

Materials and methods

Plant materials and DNA extraction

To validate the developed EST-derived markers and to screen for polymorphic ones, 13 sugarcane commercial clones, three *Saccharum* spp. (*Saccharum officinarum*, *Saccharum sinense*, and *Saccharum barberi*), and the precommercial varieties SP80-180 and SP80-4966 were used (Table 1). These plants were supplied by Centro de Tecnologia Canavieira (CTC) (Piracicaba, São Paulo, Brazil) and were chosen from a previous study of their genetic diversity using AFLP and pedigree information (Lima et al. 2002). SP80-180 and SP80-4966, not formerly used by Lima et al. (2002), are the parents of a mapping population from the CTC genetic mapping program (Garcia et al. 2006; Oliveira et al. 2007) and were incorporated in this study with the intention of evaluating the potential of the EST-based markers as segregating markers for genetic mapping. To do so, a random sample of six F_1 individuals from such population, together with both parents, was evaluated but was not used in the genetic diversity analyses. Genomic DNA from the above plants was isolated from 300 mg of lyophilized young leaf tissue using a CTAB method described by Hoisington et al. (1994) with minor modifications. DNA concentration was estimated in comparison with known concentrations of lambda DNA in 0.8% agarose gel.

Data mining and development of SSR markers

SSR markers were developed from 2005 sugarcane consensus sequences derived from the SUCEST database (http://sucest.lad.ic.unicamp.br/en/) as described in detail by Pinto et al. (2004). In short, ESTs comprising both 5′ and 3′ sequences from cDNA libraries representing 26 different tissues or developmental stages were scanned for identification of SSRs. Subsequently, a set of 342 locus-specific primers flanking SSR motifs was designed using the Primer Select software (LaserGene versions 5.01/5.02; DNAStar, Inc., Madison, Wisconsin) and synthesized by Invitrogen and Im-

Table 2. Details of the EST–SSR markers developed from the SUCEST database.

SSR[a]	Motif	Forward primer (5′–3′)	Reverse primer (5′–3′)	Melting temperature (°C)	A[b]	Foreseen size (bp)[c]
ESTA12	$(TA)_{29}$	AGGCTAGCGCTTTGAACACT	CCAAATGCCAATCCAACAC	51	11	269
ESTA14	$(AG)_{23}$	GGTGGCGGAGGGAGCAGGAT	GTCCCCGTCCCGTCGCAGAT	63	7	186
ESTA15	$(CT)_{24}$	TGAAGCAGCTAGCGGTCCAC	GTTCTCGCGGTTGATGTCCA	56	8	136
ESTA16	$(AG)_{25}$	AGTCCCTCTGGCCCTCACAC	GAGGCTCTGTGATGGGTTCG	54	10	165
ESTA17	$(AG)_{32}$	ATCCGTCCGCCGTTCCTCAG	GGCTCGCCGTTCACCATTCA	60	11	212
ESTA19	$(CG)_{9}$	CGCACCCGTTGACGAAGCAGT	GTTCCTCGCGCTCCTCTGCT	58	2	191
ESTA22	$(TA)_{30}$	AGGCTAGCGCTTTGAACACT	ACGTAAGGTGCTGGTCATCA	50	4	185
ESTA26	$(TG)_{11}$	GGCAGCCCCACATCTTCCT	GGGCACAAGCATCCGAACC	56	10	183
ESTA31	$(TC)_{9}$	AAGCTACCACACCCGATGAT	GTCGTCGTCGCCTGATGA	52	7	204
ESTA32	$(CT)_{12}$	CGTTGTACACATCCTCCTCCTCTA	TCTTGTTGATCGGGGTAGGAG	53	9	104
ESTA33	$(CT)_{7}$	CAGGGCTGGGGTAGGTTG	AGCAGACTTCCAGACCATAACC	52	8	164
ESTA34	$(CT)_{8}$	CGCTGCATCCGTCACCT	GGCTAAGAGATCGGAGAAGGTC	53	15	119
ESTA36	$(AT)_{15}$	GTTTGGGGAGAATCAGCACA	AACATGTTGCCTGGTCTGC	51	4	179
ESTA37	$(TA)_{17}$	CGTGACTGTTGCGGTTTTGT	TGGGAGCAAAGAATGAGGTG	53	3	190
ESTA38	$(TA)_{16}$	TGCTCGAAACCCCATTAGTG	GCAGATGCTCCCGTGTGG	55	5	200
ESTA44	$(CG)_{8}$	GAACCCTGACCCGAGCCCTGAG	CTCCCGCCACTTCCTATCCCTTCC	64	6	153
ESTA47	$(CA)_{9}$	AGCGACAAACTGGTGACGACT	TCTGATCTGCTGTGGGAGGAC	54	8	262
ESTA48	$(CA)_{8}$	AGCAACTCCGGCCTCTCCTG	CTTTCTGTTTTGCTCCTCCGTCTG	58	14	254
ESTA49	$(AC)_{8}$	CTTCAAGCCATCTGTCTCCTCT	TGAGGCCTGAGCAATGAAAC	53	9	167
ESTA51	$(GT)_{20}$	TCTCCCTCCTCTCGTCCTCT	CGATCTTGTTCCTTGCTGTC	50	8	267
ESTA52	$(GT)_{22}$	CGCCCGCCCGATTGCTACTT	ACCCGACCGGCCAGCAAATG	64	10	270
ESTA53	$(TG)_{8}$	TGGAAATGGCAGCTGGTCTCGT	ATGCACGTACCAGAGGGAGATTTG	58	10	186
ESTA54	$(TG)_{8}$	TGTGTCTGGTGATCTGTCGTGTG	CTCTCACAACAACCCATTCCTCTC	56	7	124
ESTA55	$(GT)_{8}$	GCGGACGTCTCATGTTCTTTG	CCCCCAAAGCAGGCAGTT	55	7	282
ESTA56	$(GT)_{7}$	ACGCTGCTTCTGCTGGTTGTCG	GCTGCCATTGTCGTTTTTCCTTCC	62	6	143
ESTA57	$(AT)_{33}$	GCAGAGCAACGAACGA	CACCCACTGAGGCGAGATA	51	3	220
ESTA58	$(TA)_{26}$	GTCGTGCCCAACATCAAT	GTGGGTCAACTCCTCTTACAT	47	13	255
ESTA61	$(AT)_{12}$	ACCTCAGTCTCCTCCTCAACC	TATACTACACATGCACAGGCTACG	52	10	232
ESTA63	$(CT)_{7}$	CGCCAGCGCTATGACGAC	GCAATTAACGCCAGACAGATACG	56	8	181
ESTA64	$(GA)_{8}$	TACAAAGCCCAGACTAAACGAC	TAGTGATGAGATGGGCAGGAAG	53	9	174
ESTA66	$(GA)_{10}$	GCTGCTGCTGGTGGACTGCT	CGCGACGCCGAAGTTGC	59	11	233
ESTA67	$(AT)_{13}$	CAGTGGATTTGCGTAGTAGAGC	ACGATGTTCATTTTGGGGGTAT	54	5	237
ESTA68	$(CG)_{7}$	ACAGTGTTGACCAGTAGGAAGAAT	CAGGTACTTGGCGGTCTTG	51	5	234
ESTA69	$(CA)_{7}$	GTCTTCTACAGGGCAAATCCAACC	TGATCTGAGTGTGGTATGTTGTTGC	56	8	206
ESTA70	$(AG)_{8}$	GATGGAACCTGAAGATGAAGAGCA	CCGGCCGGAGCACAGACG	62	11	175
ESTA71	$(TC)_{9}$	CACGCGTCCGGAGATACTG	ACTGATGAATGGAGCAAACTACTG	52	5	185
ESTA72	$(TA)_{29}$	AACTGCCCCAACGATTTCCACTG	TTGCTGCTCCATGTCGTCGTTTTC	62	9	269
ESTB12	$(GCC)_{5}$	CGCGTTCCAGATCCTCGTAT	TCAGAATCCGCAGAGTGTGG	53	6	147
ESTB13	$(ACG)_{5}$	CAGGAGCGTGAAGAAGATGG	GACCAAGTGCTTTCCATCTGAC	53	5	133
ESTB14	$(CGT)_{8}$	TGAGGGAATGAATGGACTGG	CCACCACCACCATACCTGTC	52	12	297
ESTB16	$(CGG)_{5}$	CGCGCCCGTGCTTGTTGC	ACGCATTTGGGCTGGTTCG	58	6	234
ESTB17	$(GTC)_{6}$	CTCCTCCTTGGCGGGCTCGTC	AGCAGGAGCCGCCGCAGAAG	63	6	158
ESTB20	$(CCG)_{10}$	CGATTCGGCCGTTTTCTGC	TAGTGTGGTGGGTGCTCGTC	53	3	157
ESTB23	$(CCG)_{6}$	GTCGAAGGGCGGGAAGGTCT	CGAGGCGCGGAACATCAA	58	4	103
ESTB25	$(AAT)_{7}$	GGCCTGAAATCTGAATCCTC	TCCACCCCGATGATTAGC	50	6	191
ESTB27	$(AGT)_{6}$	ACATCGAGGAGGAGGTGGAG	ACTTGACCGAGAGGCATTTC	51	6	213
ESTB35	$(CCG)_{5}$	TCGAAGGGCGGGAAGGTCTG	TGCGATGGCGAGGTCTTCC	58	6	180
ESTB37	$(GCC)_{6}$	CACCAGCACCGGCGTCCAC	GCCGACGAGGACGCTGGACA	63	6	187
ESTB38	$(GTC)_{8}$	CCACGCCAAGGAAAGGGAGAAAA	CGAGCCGGGGCGAAGACC	62	5	210
ESTB39	$(CGC)_{8}$	CGCGTGTCCCCCTCCCAAGTT	GAAGATCGTCGTGCGGCTGTCA	62	9	146
ESTB40	$(GCC)_{5}$	AAGTCCAAGGCCAAGTCCAG	GACCTGAACCGACCGACTGA	55	7	155
ESTB41	$(CGA)_{8}$	CATGGAGAGCTGGGCGACCTG	GGCGGCGGCGAGGATGA	63	21	163
ESTB43	$(CGA)_{9}$	TGGCGTCGCACATCATCCT	CGGGGAGGTAGTCGGTGAAG	56	4	330
ESTB45	$(GTA)_{7}$	ACGAAGCGGCGGTCCTC	TATTGCAGCCCACGCCTTCT	57	12	110
ESTB47	$(CGC)_{5}$	ATGATCATGCCGCTCTTGGTG	CTCCACCCGTCCCGCTACC	59	7	236
ESTB48	$(CTA)_{5}$	CTCGTCCATGGTGCCTTGTG	GCGCCCAACTTCATCCACAC	56	3	218
ESTB49	$(GCC)_{5}$	CAGTTGCCGCCGGAGTCACC	GGGCCACCAACAACGAGTCC	59	3	454
ESTB51	$(TCA)_{6}$	GTCACCAGCAACCAAGAGC	GCTGCAAGGTGACAAATCTG	51	6	416
ESTB52	$(CCG)_{5}$	GCAACCCGCCCGCTCACC	GCGGGAGCGACGATGTGGAC	62	11	145
ESTB53	$(GCG)_{5}$	ATTGTGCATGGCGGGTGAG	GCAGGAGCTTGTCGCACTTG	56	12	113
ESTB55	$(CCA)_{5}$	CTTCTTGGCCTTGGCGTTACTGA	GCTAGCTGGCCCCATTTCCTCT	60	2	116
ESTB56	$(ACC)_{7}$	TTCACGAGCCCAATGTCC	CTCGGGCTGTGGATGCT	51	7	236
ESTB57	$(ACC)_{5}$	GTGCTCGGCGACCTCTGC	GCCCCCGGACTTGTTAGC	54	7	267
ESTB58	$(CCA)_{6}$	GACGGTCCCGCTCTTGTTCTTGTA	TCCGTGAGCAAAGACAACTGACCT	59	16	179
ESTB60	$(TTG)_{10}$	AGCCGCAATGAATCCAACTG	CTCTAGCTCCGACGATGATACCTC	55	19	198

Allele range (bp)[d]	PIC	DP	Best-matched protein[e]	GI[f]	Biological process[g]	Molecular function[g]
223–251	0.87	0.98	No hit			
193–207	0.78	0.97	No hit			
118–139	0.83	0.97	Caffeoyl CoA 3-*O*-methyltransferase	32347451		1
167–189	0.87	0.99	Hypothetical protein	50900112	12	15
146–192	0.75	0.82	Hypothetical protein	34894358	14/12/17	1/7
202–216	0.21	0.30	Peroxidase 7	57635159	14/5/3	5/6
180–194	0.65	0.80	Hypothetical protein	49388878		
171–200	0.87	0.98	TPA_inf: WRKY transcription factor 74	46394402		8
229–273	0.74	0.82	DFL1	15239653		
99–117	0.84	0.97	Os03g0675300	115454569		
173–195	0.81	0.97	Hypothetical protein	33327288		15
152–127	0.90	0.99	Hypothetical protein	12597881		
156–201	0.48	0.83	No hit			
227–240	0.58	0.58	BTH-induced ERF transcriptional factor 1	56567581	14/7/8/6/12	8
174–204	0.76	0.84	No hit			
161–198	0.80	0.75	OSJNBa0086B14.26	50924658	18/14/12	15
275–324	0.77	0.94	Putative branched-chain amino acid aminotransferase protein	34904468		1
241–289	0.88	1.00	Chloroplast phytoene synthase 1	38037628		
173–188	0.87	0.97	Putative glutathione *S*-transferase	50918191		1
235–299	0.85	0.96	Putative OsNAC7 protein	6006373		
267–302	0.87	1.00	No hit			
192–233	0.84	0.97	Indeterminate spikelet 1	2944040	14/7/8/6/12	8
140–150	0.78	0.92	No hit			
288–339	0.80	0.98	No hit			
147–179	0.76	0.70	Unknown protein	48843751		
170–285	0.59	0.11	No hit			
219–313	0.88	0.99	No hit			
241–280	0.85	1.00	Lipid transfer protein-like protein	39748092		
177–209	0.83	0.99	TPA: class III peroxidase 90 precursor	55701047	14/5/3	5/6
213–349	0.83	0.97	Sulfotransferase domain containing protein	77551106		1
306–357	0.87	0.99	Putative protein kinase	50901454	14/12	1/7
240–293	0.58	0.87	Os01g0195000	115435060		
125–132	0.74	0.93	Fructose 1,6-bisphosphate aldolase	40362980	14/12/18	9
306–367	0.84	0.97	Pectatelyase 2	6606534		9
180–225	0.88	1.00	Peroxidase	51038054	14/5/3	5/6
203–366	0.72	0.90	Plastid starch synthase I precursor	51832613		1
271–342	0.85	0.90	Protease inhibitor/seed storage/LTP family protein	110289546		
107–182	0.71	0.60	Unknown protein	51969384		
132–145	0.67	0.86	Peptidase/subtilase	79318240	14/12	15
263–289	0.90	1.00	RNA-binding protein RZ-1	1395193		8
262–306	0.71	0.87	SbCBF5	60593387	14/7/8/6/12	8
94–166	0.80	0.86	BRI1-KD interacting protein 128	42733518	14/7/8/6/12	8
173–176	0.56	0.66	Hypothetical protein	56784702		8
60–81	0.69	0.71	MFS18	22647		
192–207	0.78	0.86	Putative benzothiadiazole-induced somatic embryogenesis receptor kinase 1	53792830	14/12	1/7
203–232	0.78	0.94	P0480C01.19	34910656		
199–217	0.74	0.82	MFS18	22647		
182–214	0.77	0.87	Putative *O*-methyltransferase	55297295		1
207–229	0.72	0.81	Os05g0553400	115465327		
111–144	0.83	0.96	Alpha 3 subunit of 20S proteasome	34898416	18/14/12	15
165–188	0.79	0.91	Blue copper-binding protein-like	50882442	14	14
81–165	0.93	0.98	Hypothetical protein	50916060		
255–290	0.67	0.78	Putative diphosphate-fructose-6-phosphate 1-phosphotransferase	55168177	14/12/18	1
94–160	0.84	0.99	OSJNBa0032F06.16	50929735	14/12/19	12
220–242	0.82	0.83	Putative zinc finger transcription factor	50933107		8
199–213	0.57	0.29	B0103C08-B0602B01.7	116311124		
322–373	0.59	0.60	Glycine-rich RNA-binding protein	10799202		8
339–372	0.76	0.97	Hypothetical protein	6714472		
89–149	0.87	0.95	Os02g0809900	115449437		
95–142	0.90	0.81	Os06g0524300	115468316		
120–123	0.35	0.61	Beta adaptin-like	54290350	16/2/17/1/8	
232–254	0.76	0.86	Glycine-rich RNA-binding protein-like	50934801		8
270–293	0.75	0.79	Transcription factor RAU1	19401700		
127–205	0.92	0.99	Putative myb protein	33087065		8
159–272	0.92	0.99	OJ000114_01.16	50924506	12/14	8

Table 2 (*continued*).

SSR[a]	Motif	Forward primer (5′–3′)	Reverse primer (5′–3′)	Melting temperature (°C)	A[b]	Foreseen size (bp)[c]
ESTB63	$(TTG)_7$	TCCCTGCAGCGATCAAGAACT	TCAAGGCCGAACGGATTACAA	57	10	213
ESTB64	$(TTG)_9$	CACGTGCCTGAACCTTGATTG	CCTATAGGGGTTGCACGAGTTGT	56	5	166
ESTB65	$(CAA)_5$	CGATCCCAGGCCAGTTCAT	CTTGTTCGCTACTGCTACCGTGTT	57	13	221
ESTB66	$(CAA)_5$	GAACCCTGCCGCCTACTGG	AGGTAGGCGGCACGGTTTG	57	10	150
ESTB67	$(CAA)_8$	GCGAACGCTGCTACATACCA	GACTGAGCGCGGGTAGGATA	54	13	196
ESTB68	$(CAA)_6$	CACTGGCCCTGGCGAACACTG	CCGCAGCCAGTGGGTTTGATG	62	9	153
ESTB69	$(CAA)_5$	ACGAGGACGAACACAGCACTAACA	CCTAGTCCAGAGCGGGGTTTATC	57	13	144
ESTB73	$(GGT)_5$	AGCAGGCTCGGCGTGTAG	AGACGCTGGAGATGGACACG	55	13	217
ESTB75	$(AAG)_5$	GACGGCAGGGAGGCAAGACGA	GCCCCGACGAGCCTGGACA	63	5	123
ESTB80	$(CGT)_{10}$	ATTTGGCCGAGGTGGGAAGATG	CCGACCAGGCAGAGCACCAC	60	5	114
ESTB81	$(CGT)_5$	GCAGTACAGGGGCGTGAGG	CGTCGACGTCGCTGCTCT	55	8	302
ESTB82	$(CGT)_9$	CGTCGATCGAGATGAAGAAGG	GAAGCAGTCGTGGAAGTGGAG	54	6	263
ESTB85	$(GCG)_5$	CCTGCGTCTGGGCTAATGG	GCATGCGGCGGTAACTCA	55	2	129
ESTB86	$(CGG)_5$	GTGCATTGCGCCGACGAC	GGCCCGGTACCGCTTCTTC	58	3	123
ESTB91	$(CTT)_5$	CAGCTGAAGTTCTTGCCGAGGAT	CCTCGTCGCCGCCATCC	56	2	168
ESTB92	$(TTC)_5$	TCTGAATGGATGTCGCCCTGTG	TTTGCGGGCTTCTCTGCTTTCT	59	3	241
ESTB93	$(CTT)_5$	TTCTCTGCCAAGTTATCCACCTC	CTCTCGTGCGCGTGACATC	55	6	197
ESTB94	$(CTT)_9$	GAGGCAGCCAGGCAGGTCAC	GGTGGCAGTGTTCAGGCAGATG	59	12	236
ESTB97	$(GAC)_7$	GCCACCATCGTCGGGAAGA	CGCGCCACCACCAACTACA	58	9	221
ESTB99	$(TCG)_5$	GAGGTCCTTCTTGTAGTTGTATGC	GTGCCGGAGGATTTGATG	51	3	219
ESTB100	$(TCG)_6$	CCACGGGCGAGGACGAGTA	GGGTCCTTCTTCGCCTCGTG	58	11	268
ESTB102	$(GGC)_5$	AGGTGCACTCCGTCTGGTC	GGAGCCAGTACACGGAGATG	52	6	207
ESTB104	$(GGT)_7$	AAGGCGTTTGCGTTCTACA	CTTGAGAGGGAGGAGGCTGAT	54	4	133
ESTB106	$(GGT)_5$	TCTTCCCCAGGGCATTCTCC	ATAGCATCTGGGGAGCGTGGTT	59	9	211
ESTB108	$(AAC)_7$	AGAAGCAGCGGACGATACAGATGG	GGCGGCCGGTGGTTACAAT	59	8	195
ESTB110	$(TGT)_9$	GCCCTTTATGTCCCTTTATGTGTC	TCGCCTCTAACCCCCTTGAC	56	8	190
ESTB111	$(TGT)_7$	TCATCTCTAGCCTACCCCAACT	ACCCATAAGACTTTGCGTTCAC	53	8	179
ESTB112	$(TTG)_5$	TTGTTCGCTACTGCTACCGTGTTG	CGAGGACGAACACAGCACTAACAA	58	5	173
ESTB114	$(GCG)_5$	CCGCCGTGGAGGGGAATAAAGAG	GTGAGCACGCCCGCAGAGACG	65	6	210
ESTB115	$(CCA)_5$	CCCGGCCACATCTGCGTCTTC	TGGTCGGCGCCCTGAGTGG	64	9	166
ESTB116	$(TTC)_5$	GCTGCTGGCCAAGGAGTTC	CGGTCCGGTAGCAATCCAG	55	13	186
ESTB117	$(TTC)_5$	ACGTTGGCTGTGGGTTTG	CGACAAGGACAAGGTGCTC	50	6	284
ESTB118	$(TTC)_6$	CTTGGCTAGGGTTTCTTGAGTCGT	CATGGCTTTTGGCTTGCTTCT	56	11	110
ESTB119	$(AAG)_5$	TGATGCGGCAGCTGGACGAGAC	ACGCCGAGGCCGAGGATGGT	65	6	157
ESTB122	$(CAG)_7$	CGACGGCGATGAAGAAGACCAG	GTGAGGGTGCGGAGGTGGAACT	61	8	214
ESTB125	$(CAG)_9$	CGAGATGCGCAAACAGAACAAG	CCGTAGGGGAGTCGTCGTG	55	12	179
ESTB128	$(ACC)_6$	GCAGGCGCTCATCTACAAGT	GGAGCACCGCCACTTCTT	53	3	235
ESTB129	$(GTT)_6$	GGTCCGCCGTCGTTGAGTTTGA	CCTCTCCCTCGGCGACTCCATC	63	3	156
ESTB130	$(CAA)_{13}$	GCCCAGGTAATTATCCAGACTC	GCTGTTGCTCACTGGTTCC	50	17	124
ESTB131	$(TGG)_7$	GGTGGTGCTCGATTTGGTG	TAGTTGCACCTCCACCGTAGAC	54	14	163
ESTB132	$(GCG)_{10}$	CAGGATGCAGGCACAGAGGT	ATTCGGCTGGCGCTCTATTT	56	6	250
ESTB133	$(TCG)_7$	ACGTCGTCCCCATCCTTGTCA	GGATCCGCCCAGCCCTAAC	58	14	267
ESTB134	$(GAC)_9$	TCCCTTGACGACGCCTACGA	TCGTCTGCGCTGCTGTTGTC	57	6	197
ESTB135	$(CGA)_8$	GCAACAAGGCTGAGGAGGAAGT	CGGCGAGGAAGGAGGAGAAG	58	5	228
ESTB136	$(CGA)_5$	GGCACTGAGCGGGACCACCAC	GCTGCGGCTGCGGGAGAC	61	10	162
ESTB139	$(GGT)_6$	GCACCGTTGGGCGACAGGA	CATGCATGCGTGGTGGGTGTAT	60	8	284
ESTB140	$(TGG)_6$	CCGAGGGGTAAGGGTAGT	CCGCTTCCTTTTATCTCTTC	48	14	202
ESTB141	$(AAC)_{13}$	TGACTGGAACAAAGACGGACTG	TTTCATCTCTAGCCTACCCCAACT	55	7	180
ESTB145	$(CCT)_5$	GGGAAGCAAGCGAGAGCAGCAGAG	GAGCGCGAGGCCGTTGTTGAG	68	16	234
ESTB146	$(CCT)_6$	AGCTAGCGAGGCCACCAAGTCCAG	CCCTCGCCTCGCCCACCAT	64	7	142
ESTB147	$(CAG)_{11}$	AGCAGCAGGGCGGCAACAACC	GGGCGCGACAGGGGCTTCTC	65	8	221
ESTB148	$(CGA)_6$	CCCCCTTGTTCCTGTTCCCCTCTT	GTACCCGTCGTCCTGCTCCTCATC	62	5	168
ESTB149	$(GGA)_6$	CGAGAGGAGCTACCGGTCAGTGTC	CGAGCTCGTTGCGGTTGTGG	61	15	125
ESTB150	$(AGC)_7$	GCCTTCTTCTTCATCCTTCTCT	CTCCGTGCTGTATCTTCCTCTG	53	10	139
ESTB153	$(GTC)_6$	GCTCTGCGCCAACAACTGC	CGGCGCTGAAGCTGAGGAAG	60	9	151
ESTB154	$(GAT)_5$	TTCTGCGGCAAAGGTTCC	TAGGATCTGCACCACTTGACACTT	55	8	155
ESTB155	$(ATG)_6$	ACTAAGGCCTCGGTTTGGTTTC	ATTTCCATGCCAGCACAGTAGC	61.0	5	175
ESTB157	$(GGC)_5$	TCTCGTGGACGGCTACAACATC	CGCGCCGTTGAGGTCCAC	60	8	127
ESTB158	$(GAC)_5$	CCTGCCGACGGGGAACTG	CGGTTCAATCAAATCCCTCTCCTC	59	6	163
ESTB160	$(GAC)_5$	CAGCAGCAGCAGTAGGGAAGT	AAACTACTCGCCCCCATCATC	55	13	197
ESTB161	$(ACA)_6$	TGCCTACATTACCTTTCCTCAC	GGTGATGGTAGCGAAGTAGTTG	51	2	249
ESTB162	$(CCG)_5$	GCTTGGCGTCGGTGCAGTAGTTG	GGCGTCGTTCCGTGGGTTCAG	63	2	209
ESTC13	$(TGCG)_5$	CAGCCAGCAACAGAAAGAAGAGGAG	TAGAGCGCATATCGGCAGTCATTC	53	7	356
ESTC15	$(TAAT)_{13}$	TGTTGGCGTGCTGCTCTGTG	CCATTGGCTCGGCTGTAGGA	61	5	203
ESTC16	$(TCTA)_4$	GGCTGAGCTAAACCTAAACGAC	TATCACAGAGTACACGACGACACG	61	3	229

Allele range (bp)[d]	PIC	DP	Best-matched protein[e]	GI[f]	Biological process[g]	Molecular function[g]
325–369	0.84	0.92	Mitogen-activated protein kinase 6	37594657	14/12	1/7
202–223	0.70	0.95	RNA-directed DNA polymerase HMG-I and HMG-Y, DNA binding	87162631	12/14	8/1
230–286	0.87	0.99	Putative ribosomal protein L32	50947773	14/12/19	2
104–167	0.87	0.96	Alpha kafirin	22208486		
194–261	0.90	0.97	Putative kafirin preprotein	22208459		
193–234	0.86	0.97	Putative kafirin preprotein	22208459		
148–169	0.89	0.99	Putative kafirin preprotein	22208459		
126–276	0.90	1.00	Hypothetical protein	110737787		
117–142	0.68	0.84	Hypothetical protein	34902434		
118–138	0.76	0.88	Ankyrin-like protein	50902216		1
159–271	0.83	0.97	Putative AP2/EREBP transcription factor	83658836	14/7/8/6/12	8
259–278	0.79	0.93	Putative peroxidase	34913714	14/5/3	5/6
138–145	0.25	0.37	Endotransglucosylase/hydrolase XTH5	51039064	12	15
134–152	0.59	0.55	Myb, DNA binding	92898924	14/11/19	8/1
220–228	0.33	0.64	DNA binding/protein binding	22330295	14/7/8/6/12	8
251–257	0.54	0.73	Putative sugar transporter	22208506	1	16/17
185–254	0.80	0.78	Eukaryotic translation initiation factor 2 subunit beta (P38)	3122234	15/16/12/14/19	8
249–303	0.85	0.98	No hit			
146–244	0.83	0.98	Ethylene-responsive element-binding protein homolog	4099914	14/7/8/6/12	8
237–246	0.55	0.76	No hit			
367–308	0.88	0.99	22 kDa drought-inducible protein	49234816	3	
269–298	0.65	0.31	WRKY transcription factor	44893902		8
121–138	0.63	0.78	Hypothetical protein UM03029.1	71017891	12/14/4/3	8/7
207–271	0.85	0.93	SLT1 protein, putative, expressed	108707470		
167–238	0.85	0.92	Hypothetical protein	50939565	14	
202–239	0.82	0.96	Putative polyprotein	48209910		
126–199	0.84	0.95	No hit			
176–199	0.73	0.87	Unknown protein	42562411		8
234–284	0.79	0.95	Expressed protein	77555287		
107–166	0.86	0.88	No hit			
181–210	0.87	0.97	P0497A05.15	50902306	14/7/8/6/12	4
307–335	0.78	0.91	Unknown protein	50508660	1	14
109–147	0.85	0.86	Homeobox transcription factor GNARLY1	32351469	14/7/8/6/12	8
120–184	0.80	0.99	Endo-1,4-beta-glucanase cel1	7527353	12	15
240–281	0.83	0.98	Hypothetical protein	50931511	14/12	15
224–258	0.89	0.99	Hypothetical protein	15217316		
223–254	0.57	0.80	Hypothetical protein SBB1_21t00001	113208405		
197–226	0.59	0.11	Putative GDA2 protein	57899411		
94–132	0.92	1.00	Putative auxin response factor 7a	42409013	14/7/8/6/12	8
110–156	0.84	0.90	Putative helicase	50919223		
260–281	0.75	0.91	Putative UPF0183 protein	21554302		
198–262	0.90	0.97	HSV-I stimulating-related protein-like	57899937		
183–234	0.78	0.95	Putative pectin methylesterase	51090795		
215–232	0.74	0.76	Protein disulfide isomerase	59861265	14	12
166–214	0.86	0.95	Putative AP2 domain containing protein	50920393	14/7/8/6/12	8
260–317	0.84	0.96	Indeterminate-related protein 9	55418544		
180–218	0.90	0.96	Hypothetical protein	50910633		
196–227	0.78	0.97	Hypothetical protein	19224990		1
259–382	0.92	1.00	Alpha-galactosidase-like protein	7572929	12	15
134–192	0.75	0.96	OSJNBa0071I13.13	50929299	14/5/3	5/6
—	0.76	0.87	Putative ATP-dependent Clp protease ATP-binding subunit ClpX1	47847699	13/9/1	7/15
122–201	0.76	0.82	Putative beta-amylase	110742879	12/18	15
105–149	0.92	1.00	Cellulose synthase-7	9622886	12/14/19	1
135–170	0.85	1.00	OSJNBa0011P19.5	34894718	14	1/7
156–192	0.84	0.93	Multiple stress-associated zinc finger protein	37548823		
150–180	0.84	0.88	HD2 type histone deacetylase HDA106	14550106		
156–205	0.72	0.86	OSJNBa0043L09.23	50928999	12/14	8
110–162	0.84	0.86	Putative thaumatin-like protein	37531042		
140–197	0.71	0.72	Histone H2A.4	75282487	15/16/12/14	8
168–256	0.86	0.99	cAMP response element binding protein	92879198	14/7/8/6/12	8
190–233	0.35	0.53	Lipid transfer protein	1041815	1	10
244–251	0.25	0.37	Endotransglucosylase/hydrolase XTH5	51039064	12	15
303–340	0.79	0.97	Putative actin-depolymerizing factor	34897072		4
152–215	0.71	0.84	No hit			
227–251	0.45	0.31	Phospholipid transfer protein 1	75994570	1	10

Table 2 (*continued*).

SSR[a]	Motif	Forward primer (5′–3′)	Reverse primer (5′–3′)	Melting temperature (°C)	A[b]	Foreseen size (bp)[c]
ESTC17	$(ACGC)_5$	TAGATCAACCACGCTTCACG	TGATGAGCTAGCTGGCAACA	61	15	137
ESTC18	$(AGGA)_3$	GGCAAGAACACGCGAACACGAG	AGCCTCCCTGCCTTTTCCCACTC	63	4	129
ESTC19	$(AGCC)_5$	TGCTGGAGTGGAGTGCTACC	GACGACGACGAATCCATCAG	61	8	238
ESTC21	$(TAGA)_4$	CGACACGTACGTACATGCGGAGAG	CGTCCCTTACACCATCAGCACCTC	59	7	226
ESTC22	$(CTCG)_3$	CTCCGTTTCTCAGCTCCCTCTC	AGGAAATGGTGCTGCTACGAAC	61	13	171
ESTC23	$(CTTC)_3$	GCCAGAAGAAGAAGCAGGAGAACG	TCACCGTTCAGCTCCTTCCACTTC	59	4	172
ESTC24	$(AGGA)_9$	TCCTCCCCCTCCTCTCCTG	GCAGTCGGGGCAGAAGGA	63	9	175
ESTC25	$(ATTC)_4$	TGCTGCTGCCCTGCTACGATAC	GGCGGCAACACCTCCTACACA	61	7	253
ESTC26	$(CACC)_3$	GTTCCCATCTCCACCCACCCACTC	CGGCGGCCACGACCTTCTC	64	6	170
ESTC27	$(ATCC)_4$	TCGTCTCCGCCCGTTTTC	TCCCTCGAAACGGTCTGCTA	56	6	191
ESTC29	$(TTGC)_3$	GATTACCGGCAACCACTCTCC	CTTTTGCACGGAACCAGAGC	64	3	310
ESTC30	$(CAAG)_3$	CCCTGACTCCTCCTGACAAAAG	CAGCTGCATGCGGACCTT	59	5	161
ESTC33	$(ATAC)_3$	GCTCTCGCGCATCCATCTGAAATC	AAACCCGCGGCCCTGACGAC	64	9	129
ESTC34	$(GTAC)_4$	CTCGCCGGCAGACCACCAC	CCGCCGGCAGGAGAACCA	61	7	132
ESTC35	$(GAGG)_5$	CGTGTCGCGGGAATCGTT	AGATGCTGCAGGCGTCGTC	61	12	181
ESTC36	$(GATC)_4$	CGAATAGGACAGGGGAGGAAGAGG	CGGCGGCGGTTCTTGACG	61	8	236
ESTC38	$(TTGT)_4$	TTACCCGCGGAAAGTGAG	CGTCCTGCTCGCCATTC	61	7	142
ESTC39	$(GGAC)_3$	CGCGCCGGAGGTGTTTGT	GGCGCATGTCCCGTCACC	65	5	220
ESTC40	$(CCGT)_4$	GGCCGCTGCTTTCGCTGTCG	CGTCCAACCACACCCACCCTCAC	59	7	206
ESTC41	$(CTGT)_3$	CGCTCGCCGAACCTAATGTGG	TTCATGCAAGAGCCGGTCACAGAT	59	3	224
ESTC42	$(GACA)_6$	CGCAGCAATGGCCAGGTAAATCA	GCAAGCGTGATCGTGGTGGGTTTA	63	5	162
ESTC44	$(CAAT)_3$	CATGGTGCTGTGGGGTTCTG	GGTGGGCAAAGCTGTCCTGA	63	4	90
ESTC45	$(ATTG)_5$	GCCGCCGTCGCTGGATTG	GATGGATCCCCGCCTACCCTACAC	63	20	131
ESTC47	$(ACCT)_4$	TACATCCTCGGCCCAACTAATCAG	GAAAATTAGGGCTGGATGGTGCTC	59	8	148
ESTC48	$(TGGC)_5$	CCGCCGGACGACGACGAA	CCGGCCGGCCAGGTTAGGAT	61	5	200
ESTC49	$(GGAT)_4$	AAGGAGTGGGAACGGCGGATACG	GCCAGCCCCCTTGCCTTCCTC	61	4	190
ESTC50	$(TTTC)_4$	TCTGGGGCCATGGAAGTTGA	CAAGATGGTGACCCCGCAACTA	56	4	215
ESTC51	$(CAAA)_4$	TACTTTTGCTCCTTGCTCCACTTG	TCCGCACCAGACAGACTATCAG	59	5	124
ESTC52	$(TTGT)_4$	TTACCCGCGGAAAGTGAGC	GCGGATGTCCTTATTGCTGTTTAC	59	8	197
ESTC53	$(GGAC)_3$	GCGCCGGAGGTGTTTGTG	GCGCATGTCCCGTCACCT	64	8	218
ESTC54	$(CGTC)_3$	CGCACCCCCATCCATCCTCTC	AATCGTCGCCTCTCCTCACCTCAC	61	11	143
ESTC55	$(CTGT)_4$	AAAACCCACTACAACAGCATCC	TACGCACTTACCACAGAGCA	61	9	173
ESTC57	$(GAGG)_3$	GAGTAACTGTGTCGCCTTCA	AAGTTTGAGGCTGTTGTGC	53	8	273
ESTC58	$(GAGC)_3$	TGGGCACCGTCATATCTAAATC	GACCCTCACCCCACTCTCC	59	4	207
ESTC60	$(CTTC)_3$	CGTTAAATCCGCTCGCATACTC	TCGACGATTGGGTTGGAGAC	59	9	159
ESTC61	$(AATT)_3$	TTACCGAAGCAGCAGCAATCC	TGCCATGCCCCCAGACT	61	5	162
ESTC62	$(AGGC)_5$	AGCCCCTCACCGCCACTA	TGTACACCTCGCCGTCCAG	63	7	205
ESTC63	$(ATTG)_3$	ACTCTTAGGTAACTGCTGGTGTCT	ATCGGCATGGTATCGTTCAA	50	3	223
ESTC64	$(GAGC)_3$	CGTTCCGGGTACCACTCCTG	GCCGCAGGGGTAGCATCAT	61	4	217
ESTC65	$(ACTG)_3$	GCCCAAATCTGCCCACTCATCAA	GGTGCCGGTGGGTTCGTCA	63	5	176
ESTC66	$(CCGC)_3$	AGTACAGGCTGCTCTCAATCAA	TCTGTCATCTGTGTTCGTTCTG	61	22	152
ESTC67	$(AGCC)_4$	GAGCGAGCCAGTGAGGTCT	CCCTCGTCTCGTCCTCCTT	59	10	157
ESTC68	$(ATCA)_3$	CGGACACGTCGCTCATTG	AGTTGGCTCGGAGTTTGACG	59	6	147
ESTC69	$(AAAC)_4$	GAACGACGAGCAAGGGAAGGAATG	CAATAAGGCAGGACGGAACAGATG	56	3	159
ESTC70	$(CTTT)_3$	CGCCTCGATCCTCTCCA	TTAGCGAAGTCCATCAATCACAC	56	4	151
ESTC72	$(CAGC)_4$	CTCCCTTAAATCCCTCCTCTCC	AGAATTCGAATCCTTGGCTGAC	59	6	174
ESTC74	$(GCCC)_3$	CCGGACGCTGCTGGTGT	CCCACGGCTTGACGAGGAT	64	2	134
ESTC75	$(CTGG)_3$	CAAGTATAAACGTGCAGTAACCAG	GCGCTAGAAGGGCAACTG	61	2	210
ESTC77	$(CCAT)_4$	GCAGTGGGACGGGTGAC	GCGTCTTGTTGCTGCTGTC	64	13	162
ESTC78	$(ACCC)_3$	GTGCGGCTTCTTCCTCCCCTTCC	GACGCCCACCACCGCCATTG	63	5	163
ESTC79	$(ATGT)_3$	GTCCCCGGCCTTCTGTAACTAT	CAAACAGGAAGGAAAAGGAAGC	59	3	215
ESTC80	$(ATTC)_3$	ATTCTTTCTCCCCCTGTTGTGC	GTCGCCAGATCGCTTTCGTT	59	16	189
ESTC81	$(CCAC)_3$	TGACGATCCAGAACCAGAAGC	CGTACACAACCGCCTTCATCC	61	11	152
ESTC82	$(TGGT)_4$	GGCGGCGGCTGGCTGGAT	GATTTGTGGCTGGCGGAAGTGGAC	63	7	149
ESTC83	$(CAAC)_4$	ATTTGTGGCTGGCGGAGGTGGAC	GGCTGGCGGCTGGCTGGAT	65	12	149
ESTC84	$(GCCT)_4$	AAGCCGGGTTCCAGTCCAG	GCAACCAAAAGGCTCAGAACAG	59	10	204
ESTC85	$(GGCC)_3$	GCAGCAGCGCCGAGAAGAGCA	CGCCGGATCACCAGCCTACAGC	65	4	127
ESTC86	$(ACAA)_3$	TATGGAAGCAACGGACAGTG	CAAGCCCTGAAGGTAGTCCA	50	3	145
ESTC87	$(CTTT)_3$	CGCCGCCGCCACCATTTT	ACGGCGGCGGGGGAAGAA	56	6	208
ESTC91	$(GCCG)_3$	CGAGGACGATGTGGGAGAGG	CTCACCTCCCCCAACACAGTC	61	7	259
ESTC94	$(GCCG)_3$	CCGGCGGCAAGGAAAGCACA	TCGAAGAGGCAGGTCATCAAGA	61	4	117
ESTC95	$(CCAG)_3$	GTGCTCAAACAATGGGTAATCAA	GCTTTGCCGCTTCCAGTTTCAC	59	4	129
ESTC96	$(TCCA)_3$	TTCTTGTGCTCCTTGGCTTGCTCA	GGCTGCGGCTGCTGATGGTC	61	8	217
ESTC97	$(CAGG)_3$	GCTGCTCCGCTCTGCCTCTGT	CCGCCGGCCTTTCTTGTTTCT	61	5	179
ESTC98	$(CACG)_3$	TCGATGTGACGGGAAACCTGAAA	ACCAACGGGAGACGCCATTACAAAC	59	2	162

Allele range (bp)[d]	PIC	DP	Best-matched protein[e]	GI[f]	Biological process[g]	Molecular function[g]
115–150	0.89	1.00	Putative photosystem II core complex proteins psbY	42407913		
121–132	0.53	0.65	Root uracil permease 1	89330191	1	
242–281	0.78	0.94	P0518C01.24	34906022		
313–336	0.77	0.90	Lipid transfer protein	311331	1	10
209–259	0.79	0.97	Hypothetical protein	50939961		
300–320	0.69	0.82	Putative embryogenic callus protein 98b	47497872	14/7/8/6/12	8
155–180	0.81	0.95	General transcription factor TFIIB	18481632	14/7/8/6/12/15/16	8
296–345	0.73	0.75	CAA303719.1 protein	5777631		
192–221	0.77	0.96	TaWIN2	9798605		4
188–229	0.79	0.97	Hypoxia-induced protein conserved region containing protein	77548353		
350–372	0.43	0.63	Chain A, C-terminal domain of mouse brain tubby protein-like	56201923		
153–161	0.66	0.88	Mitochondrial uncoupling protein 4	51860691	1	
109–165	0.83	0.92	Elicitor-inducible protein EIG-J7	40287496		
130–176	0.81	0.93	Putative heavy-metal-associated domain-containing protein	48995219	1	14
183–236	0.89	0.99	Putative RING-H2 zinc finger protein	32493110		
227–254	0.83	0.94	Type IIB calcium ATPase	18483249	1	7/15/13/16
134–165	0.84	0.63	Isopentenyl pyrophosphate isomerase	33340598	14/19	12
167–227	0.63	0.65	Putative glycosyltransferase	50939609	12/14/19	1
222–270	0.79	0.81	Unknown protein	15231121		
237–256	0.47	0.29	No hit			
186–241	0.72	0.72	Putative receptor-like kinase Xa21-binding protein 3	52353644		1
187–205	0.61	0.86	B1066G12.15	34904540		
106–168	0.93	1.00	SbPCL1	71067066		8
153–168	0.82	0.80	Putative acetyl-CoA synthetase	26451642		
175–218	0.75	0.85	P0518C01.24	34906022		
175–232	0.68	0.73	mLIP15	1060935	14/7/8/6/12	8
212–234	0.65	0.84	At1g68060/T23K23_9	23463063		
88–102	0.71	0.11	Nodulin-like	92896651		
163–203	0.83	0.84	Isopentenyl pyrophosphate isomerase	33340598	14/19	12
157–212	0.79	0.97	Transferase, transferring glycosyl groups	15230679	12/14/19	1
106–169	0.86	0.90	No hit			
180–214	0.82	0.91	IBS1, ATP-binding/protein-tyrosine kinase	79346260	14/12	1/7
294–398	0.85	0.98	OSJNBa0072K14.5	50923903	14/18/12	1
227–258	0.66	0.82	Methylmalonate semi-aldehyde dehydrogenase	50934827		6
157–186	0.86	0.96	NADP-specific isocitrate dehydrogenase	34911932		6
162–187	0.71	0.84	Short-chain dehydrogenase/reductase protein-like	49388250	14/12	1/7
170–225	0.81	0.93	Glyceraldehyde-3-phosphate dehydrogenase	474408		6
189–217	0.58	0.88	NADH dehydrogenase subunit 5	25057394	14	6
197–255	0.65	0.73	Putative sex determination protein tasselseed 2	22296320		6
180–223	0.74	0.93	Galactosyl transferase GMA12/MNN10 family, putative	77551427		1
102–265	0.94	0.91	Probable xyloglucan endotransglucosylase	38605539	12	15/1
189–286	0.85	0.99	Beta-amylase	4321978	12/18	15
131–199	0.78	0.87	Endoglucanase 7 precursor (OsGLU10)	75225384	12	15
155–162	0.59	0.17	Cellulose synthase 6	9622884	12/14/19	1
151–176	0.69	0.72	Chalcone synthase 8	17978544	19	1
133–219	0.74	0.93	Serine palmitpoyltransferase subunit 2	109809703	19	8/1
233–241	0.34	0.52	Transferase	15228696		1
259–271	0.16	0.21	Arginine-tRNA-protein transferase 1 homolog	51971060		1
105–221	0.90	0.98	Putative caffeoyl-CoA *O*-methyltransferase 1	50947279		1
168–185	0.71	0.93	SNO glutamine amidotransferase	92871881		1
170–216	0.56	0.66	Unknown protein	15219915		1
219–277	0.74	0.99	Glycosyltransferase			1
118–187	0.89	0.96	Similar to ubiquinol cytochrome *c* reductase			
150–163	0.82	0.96	Cytochrome *b*5	414705		
154–253	0.90	1.00	Cytochrome *b*5	414705		
292–371	0.87	0.95	Putative acetyl-CoA C-acyltransferase	47848479		1
111–147	0.70	0.61	Putative *O*-deacetylbaccatin III-10-0-acetyltransferase	18997236		1
265–280	0.55	0.78	Inositol polyphosphate related phosphatase	92869611		15
—	0.76	0.68	Stress-tolerance zinc finger protein	112819496		
245–304	0.82	0.94	Putative fructose 1,6-bisphosphatase	50931573	12	15
128–149	0.41	0.61	Glutathione *S*-transferase GST 12	11385475		1
146–190	0.65	0.39	Inorganic pyrophosphatase	3885882		15
218–234	0.83	0.96	Putative nitrate transporter NRT1–5	34394552	1	
188–216	0.69	0.84	Putative potential cadmium/zinc-transporting ATPase 4	24060056	1	7/15/13/16/14
187–238	0.37	0.50	Inorganic phosphate transporter 3	65335879	1	13/16

Table 2 (*concluded*).

SSR[a]	Motif	Forward primer (5′–3′)	Reverse primer (5′–3′)	Melting temperature (°C)	A[b]	Foreseen size (bp)[c]
ESTC99	$(CGAT)_3$	GTTGACCGGACGAGCCAGAG	TTGCTGCTCCTCTCCTAATCGTCA	64	5	174
ESTC100	$(CTAG)_4$	CAATGTGGCTGAGGCGTAAGAT	GGTACATACATCGGTCCAGAACTCA	59	2	160
ESTC102	$(CCAC)_4$	ATCCCTCCTCGGCTCCTCA	ATCTCAAATAAGGTCTGCCACATA	59	4	257
ESTC105	$(GCCT)_3$	GCCACAGCCAGAACGAAAATC	GAGGCGAAGATGTGGGGTGAGAACT	61	7	189
ESTC106	$(TTCA)_3$	GATTATTGGAAGCTTTGGTTTTATG	GCGGCGCCGACAAGAGG	50	6	174
ESTC107	$(ATGA)_3$	GCGGCGCCGAGAAGAGGAA	ACGGACATTGAATTTGGTTGC	54	2	136
ESTC109	$(TGCC)_4$	CAGCGCTGCGACAAGTTCC	GGGTGGAGTGGAGTGGGTAAGC	65	8	165
ESTC110	$(AAAG)_3$	ACATGATCGCCGTCCTCTG	GCAAAGGCAGAAAAAGGTGTT	59	8	138
ESTC113	$(ACAT)_3$	CGTCCGCTAGGGCCGATTCCACTCT	GACTAATAGCTGAAGCCGACCAACT	64	16	288
ESTC114	$(CGTA)_3$	CCACCGGACAGGGACGAC	ACAGCAACACCACAACGGAGTA	63	5	104
ESTC115	$(TTTG)_3$	AGATCCCGGAAACGACTGG	GATGCCTGATTCCTCCTCTACC	61	8	165
ESTC116	$(TTTG)_3$	AGATCCCGGAAACGACTGG	GATGCCTGATTACTCCTCTTTCTG	61	5	189
ESTC117	$(GGAC)_3$	GGGAGCGACGAACTGACG	GATCCCGTCGCCAACAAC	61	4	295
ESTC118	$(GTCC)_4$	TGGAGGCAGCAATCAATGG	CACGTGATCGAGATGCTGAAC	61	3	257
ESTC119	$(AAGC)_4$	GGAATTAAGCTTTGCCGACACCAC	GGCAGCACCTCCCCTTCACC	63	24	159
ESTC120	$(TCAA)_3$	AGCGCGGAAAGATGGACTC	CGCTGAGCACGCCTGATT	54	8	133
ESTC122	$(CTAC)_3$	TTCTGGTGTTCTTTTCCCTCTTCC	GCAGCGCCAGCACATCTTG	50	5	287
ESTC123	$(CATC)_3$	GGTCCCTCCTGAGCCCTGTC	GGAGTGGGAACGGCGGATAC	61	10	152
ESTC124	$(TCCC)_3$	GCTGAGGCTGATGTCGTTGATG	CCGCAAAAAGGGGTAGGAGAG	59	9	135
ESTC125	$(GGAG)_3$	GCTCTCCCCCGTCCCCAAATC	GGCCGCCTCCCCTTCTTCC	63	11	108
ESTC126	$(GCTC)_4$	GCCGTTTCCTCTTCTGCTCCACTG	CTCCCTCGGTCTTTCCCCCTGTC	61	2	139
ESTC128	$(CTTG)_3$	TCAGACGCATGCACAACGAT	ACGCTACCCAAAAGAAGTCACG	54	6	135
ESTC129	$(GGAA)_3$	GGAGCAGGGTGAAGGGGAGGTG	GCCACAAGCGCGACGAGGAG	63	10	203
ESTC130	$(CTAT)_3$	CGGCAACTGGAGGAACTGAT	AACCAAACGGGCCAGTAACA	54	8	168
ESTC131	$(CTAT)_3$	CCGGCGGCAACTGGAGGAAC	AGCGGGAGTAACCAAAGCGGACAG	61	9	183
ESTC132	$(CTAT)_3$	GGCAACTGGAGGAACTGATGTG	TCGAACCAAAGCGGACAGTAAC	56	9	178
ESTC133	$(CTAT)_3$	GTGACTCCGGCGGCAACTGG	AGCGGAAGTAACCAAAGCGGACAG	61	5	182
ESTC134	$(CTAT)_3$	CCGGCGGCAACTGGAGGAAC	GCGGGAGTACCAAAGCGGACCAG	59	6	181
ESTC135	$(CTAT)_3$	GGCAACTGGAGGAACTGATG	ACACCACCAAATCCAGAACG	56	3	134

[a]ESTA, ESTB and ESTC are di-, tri-, and tetranucleotide primer pairs, respectively.
[b]Number of alleles.
[c]Size foreseen by the Primer Select software (LaserGene versions 5.01/5.02; DNAStar, Inc., Madison, Wisconsin).
[d]Estimated by linear regression.
[e]EST homology: annotation of the best homology identified by BLASTX available in the SUCEST database.
[f]GenBank identification.
[g]Biological process and molecular function categories assigned to the cluster used to develop the respective EST–SSR. Category numbers are as used in

printing (São Paulo, Brazil). The stringency criteria used were the same as described by Cordeiro et al. (2001). The SSR markers developed are referred to as EST–SSRs, since they were based on EST sequences. To explore the potential utility of the EST–SSR markers for use in research of the sugarcane structural genome, they were compared with those in the GenBank database using BLASTX directly from the SUCEST database.

PCR conditions and separation of EST–SSRs

PCR was carried out in 20 μL reactions consisting of a 10× PCR buffer (10 mmol/L Tris–HCl and 50 mmol/L KCl) including 2.0 mmol/L $MgCl_2$, 100 μmol/L of each dNTP, 0.2 μmol/L of each forward and reverse primer, 0.5 U of *Taq* DNA polymerase (Invitrogen, São Paulo, Brazil), and 40 ng of template DNA. All fragments were amplified using the following PCR profile: an initial denaturing step of 3 min at 94 °C followed by 31 cycles with denaturation at 94 °C for 1 min, specific annealing temperature for each primer pair for 1 min, and extension at 72 °C for 1 min, respectively. After 31 cycles, a final extension step was performed at 72 °C for 2 min. PCR amplifications were performed in a PTC-100 thermocycler (MJ Research, Watertown, Massachusetts). Amplification products were separated by electrophoresis on 6% denaturing polyacrilamide gels in 1× TBE buffer. To do this, PCRs were mixed with equal volumes of loading buffer (formamide containing 0.8 mmol/L EDTA and traces of bromophenol blue and xylene cyanol), denatured at 90° for 3 min, and snap-cooled on ice. Afterwards, samples were loaded on a preheated S2001 model (Life Technologies) and were run at 75 W for 1 up to 3 h depending on the fragment sizes to be separated. The fragments were visualized by silver staining according to Creste et al. (2001).

Polymorphism analysis

Owing to the highly polyploid sugarcane genome, it is often difficult to distinguish alleles from homoeologous chromosomes, i.e., to determine heterozygosity or homozygosity at a particular locus. Thus, in this statistical analysis, EST–SSRs were treated as dominant markers. The amplified fragments were scored in a binary format, with the presence of a band scored as 1 and its absence scored as 0, for each of the 18 sugarcane varieties. Both PIC and DP were obtained for polymorphism analysis of each locus. PIC values were calculated based on the formula

Allele range (bp)[d]	PIC	DP	Best-matched protein[e]	GI[f]	Biological process[g]	Molecular function[g]
157–385	0.72	0.73	Homeodomain leucine zipper protein	5006849	14/7/8/6/12	8
193–198	0.28	0.42	SOL1, transcription factor	18403510		
207–356	0.65	0.80	Putative MAPK	83320489	14/12	1/3/7
188–228	0.72	0.69	Putative protein serine/threonine kinase	63021410	14/12	1/7
198–268	0.76	0.88	Wound-induced basic protein	1172597		
133–138	0.37	0.47	Wound-induced basic protein	1172597		
173–228	0.80	0.97	Hypothetical protein	34903888	14	6
109–240	0.82	0.93	Putative cytochrome p450	115477665	14	6
—	0.89	0.97	Putative thaumatin-like protein	50726592	14/12/19	11/7
102–213	0.73	0.53	Putative phospholipase	77548280	14	15
173–219	0.84	0.98	U2AF small subunit	68036691		8
107–159	0.74	0.66	U2AF small subunit	68036691		8
276–295	0.65	0.81	Chlorophyll *a*/*b*-binding protein	3126854	14/10	
237–265	0.57	0.54	Glycolate oxidase/oxidoreductase	15231789	14	6
84–326	0.94	0.99	Putative leucine-rich repeat transmembrane protein kinase	50905839	14/12	1/7
201–220	0.81	0.98	Putative uracil phosphoribosyltransferase	50906841	19	1/7
205–290	0.74	0.83	Phosphatidylinositol 3- and 4-kinase, putative	62733536		1
175–224	0.86	0.99	mLIP15	1060935	14/7/8/6/12	8
113–237	0.78	0.91	Spectrin repeat	92896830		
90–126	0.87	0.98	RNA polymerase II transcription factor	15228267	14/7/8/6/12/15/16	8
147–158	0.34	0.53	Glycerophosphodiester phosphodiesterase	15241502	14	15
113–160	0.76	0.92	Chlorophyll *a*/*b*-binding apoprotein CP24 precursor	733458	14/10	
202–286	0.87	1.00	Putative leucine-rich repeat transmembrane protein kinase	50905839	14/12	1/7
167–200	0.83	0.93	Glycine-rich RNA-binding protein	10799202		8
206–231	0.82	0.97	Glycine-rich RNA-binding protein	10799202		8
184–218	0.81	0.97	Glycine-rich RNA-binding protein	10799202		8
308–384	0.74	0.83	Glycine-rich protein	2196542		8
91–156	0.80	0.92	Glycine-rich RNA-binding protein	10799202		8
157–159	0.56	0.64	Glycine-rich RNA-binding protein	10799202		8

Fig. 1. When more than one category was assigned to the same cluster, they are shown with "/" between them.

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where p_i is the frequency of the ith allele and the summation extends over n alleles (Mateescu et al. 2005). DP values for the kth primer were calculated based on the formula

$$\mathrm{DP}_k = 1 - \sum_{j=1}^{I} p_j \frac{Np_j - 1}{N - 1}$$

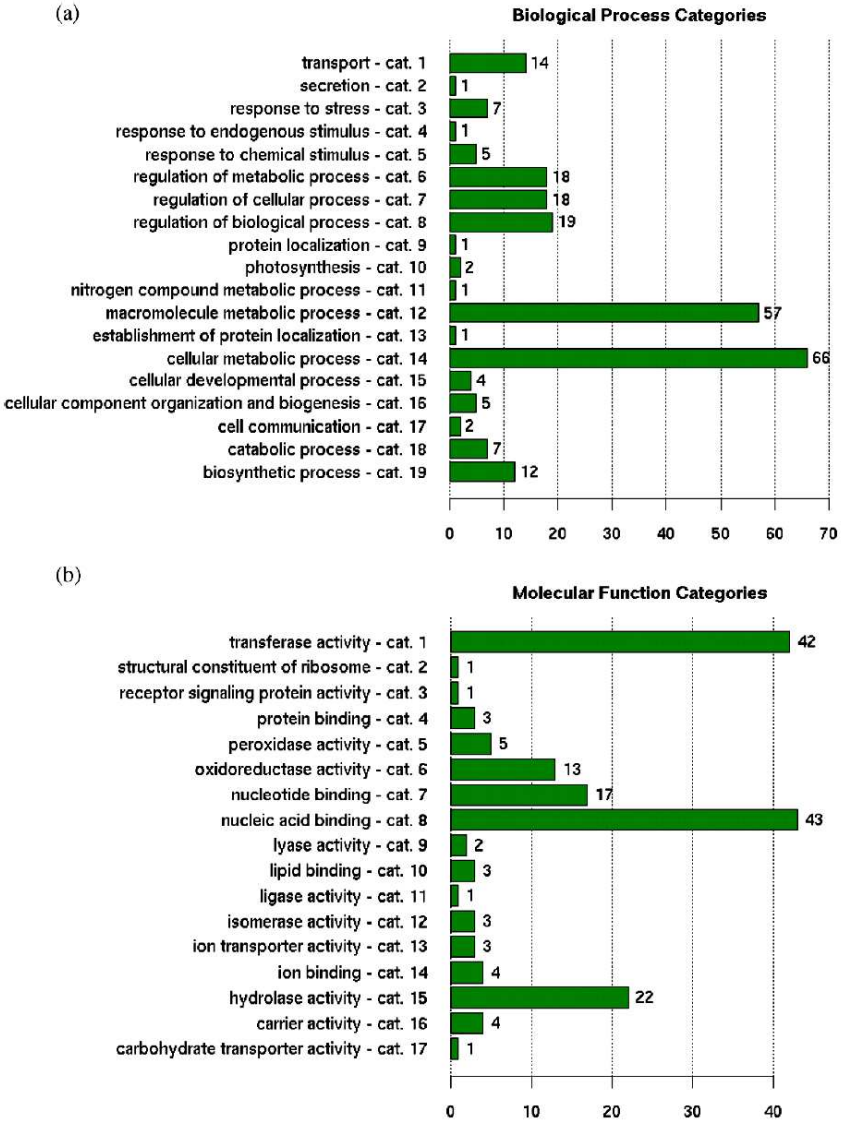
where N is the number of individuals and p_j is the frequency of the jth pattern (Tessier et al. 1999). PIC was used as a tool to measure the information of a given marker locus for the pool of genotypes, while DP was used as a quantity tool that measures the efficiency of a primer for the purpose of identification of varieties, i.e., the probability that two randomly chosen individuals have different patterns.

Genetic similarity

For genetic similarity estimate between all of the genotypes evaluated, only polymorphic bands were used in the construction of a binary matrix, reflecting the presence or absence of the fragments obtained by EST–SSRs in the different genotypes. The estimate of EST–SSR-based genetic similarity (SSR-GS) among all of the genotypes was calculated according to the Jaccard similarity coefficient (Jaccard 1901). The matrix of similarity was used to draw a dendrogram based on the algorithm named unweighted pair group method using arithmetic average (UPGMA), as suggested by Sneath and Sokal (1973).

The bootstrap procedure (Efron and Tibshirani 1993) was used to verify if the number of polymorphic EST–SSRs, used for genetic similarity estimation, was large enough to supply precise estimates of SSR-GS among the genotypes (King et al. 1993; Tivang et al. 1994). The polymorphic markers were submitted to sampling with replacement of markers to create new samples of the original data. This was done simulating different sample sizes. Genetic similarities for each of these subsets were calculated obtaining 1000 bootstraps estimates of SSR-GS for each pair of genotypes. The mean, variance, and coefficient of variation were estimated for each one of these combinations. The coefficients of variation were used to construct box plots for each sample size. Lima et al. (2002) used the same procedure described above to analyze their AFLP marker data, which

Fig. 1. Number of sugarcane clusters, among the 224 used for developing the EST–SSR markers, that were assigned to gene ontology terms from (*a*) "biological process" ontology and (*b*) "molecular function" ontology. Each gene ontology term two levels below the "biological process" or "molecular function" main branch derives one category that also receives a category number for easier referencing from Table 2.



was also used in this study to compare the results. However, the SP80-180 and SP80-4966 precommercial varieties, used in this paper, were not used by Lima et al. (2002). Thus, the AFLP-based genetic similarity (AFLP-GS) was calculated only for 16 genotypes. All analyses were performed using R, which is a language and environment for statistical computing and graphics (Ihaka and Gentleman 1996).

The coefficient of parentage (*f*) between two genotypes corresponds to the probability that alleles in a locus are identical by descent to alleles in the same locus in another cultivar (Kempthorne 1957); *f* was calculated using the "kinship package" (Atkinson and Therneau 2007). This package works within R and can be obtained at http://cran.r-project.org/web/packages/kinship/index.html. In general, the assumptions suggested by Cox et al. (1985) were adopted, and *f* was considered to be 0 among the remote ancestors.

The dendrograms obtained using SSR-GS, AFLP-GS, and *f* were compared using visual inspection. Dispersion plots and Pearson's coefficient of correlation (*r*) were used to determine the correlation level between SSR-GS, AFLP-GS, and *f*.

Results

Development of potentially functional SSR markers

Among the 43 000 consensus sequences of the SUCEST

Table 3. EST–SSR-based genetic similarity among all of the genotypes calculated according to the Jaccard similarity coefficient.

Genotype		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
CB36-24	1	—																
CB40-77	2	0.54	—															
SP79-1011	3	0.53	0.53	—														
SP70-1284	4	0.53	0.55	0.51	—													
SP71-1406	5	0.52	0.51	0.56	0.52	—												
SP80-3280	6	0.49	0.49	0.50	0.44	0.46	—											
SP79-6134	7	0.46	0.46	0.45	0.45	0.43	0.50	—										
SP79-2312	8	0.51	0.45	0.52	0.50	0.47	0.46	0.45	—									
SP80-4966	9	0.47	0.47	0.52	0.45	0.54	0.50	0.44	0.49	—								
SP80-180	10	0.51	0.49	0.52	0.51	0.47	0.46	0.50	0.47	0.46	—							
IAC51-205	11	0.50	0.48	0.52	0.47	0.48	0.51	0.50	0.46	0.46	0.51	—						
IAC64-257	12	0.51	0.46	0.48	0.49	0.47	0.45	0.44	0.49	0.47	0.50	0.47	—					
RB739359	13	0.53	0.54	0.54	0.53	0.52	0.48	0.48	0.49	0.49	0.50	0.47	0.53	—				
RB855035	14	0.52	0.51	0.55	0.52	0.50	0.50	0.44	0.51	0.52	0.50	0.48	0.48	0.51	—			
RB855536	15	0.54	0.51	0.53	0.49	0.50	0.51	0.50	0.54	0.50	0.54	0.52	0.50	0.54	0.53	—		
Gandacheni	16	0.41	0.43	0.43	0.42	0.45	0.42	0.42	0.41	0.42	0.43	0.43	0.41	0.45	0.41	0.45	—	
II76-314	17	0.43	0.43	0.40	0.43	0.42	0.43	0.45	0.43	0.39	0.44	0.43	0.43	0.46	0.39	0.43	0.40	—
Maneria	18	0.49	0.56	0.50	0.46	0.50	0.46	0.46	0.48	0.52	0.47	0.49	0.46	0.49	0.50	0.51	0.45	0.42

database used for the search for SSRs, 2005 sequences were considered useful, as they contained either dinucleotide, trinucleotide, or tetranucleotide repeats achieved as described by Pinto et al. (2004). Oligonucleotide primer pairs could be successfully designed, synthesized, and tested for 342 of these ESTs. Sixty-two (18.2%) out of this total were derived from di-, 155 (45.3%) from tri-, and 125 (36.5%) from tetranucleotide motifs. Sequences were selected based on the size of individual repeat units and on expected homology to known genes. The EST–SSRs were initially tested in agarose gels to verify the presence and quality of product amplification of predicted size. Afterwards, selected EST–SSRs were screened in 6% denaturing polyacrilamide gels for banding pattern of high quality. The details of primer pair sequences and foreseen product size with SSR motifs for the selected markers are described in Table 2.

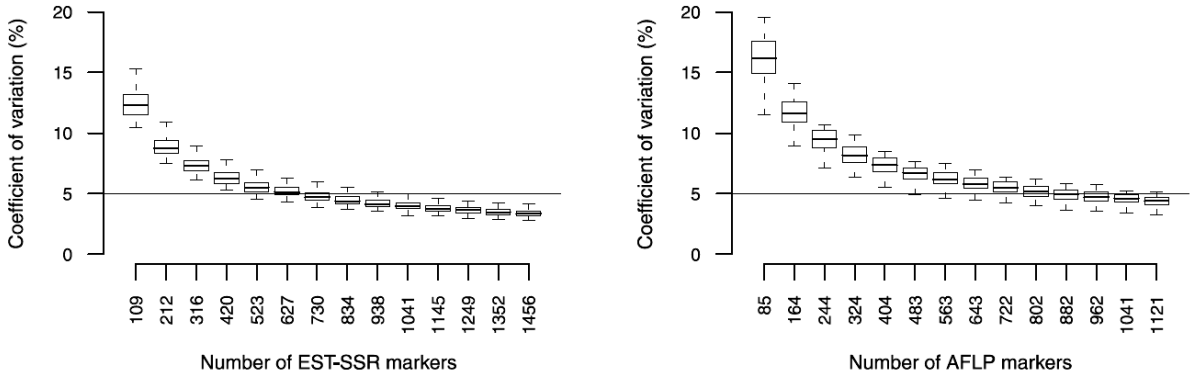
Putative functions of the products of ESTs containing SSR

To explore the potential utility of the EST–SSR markers for use in research of the sugarcane structural genome, 224 EST sequences were compared with those in the GenBank database using BLASTX directly from the SUCEST database. Out of all clusters, 82.1% showed homology to a vast range of annotated proteins including some enzymes related to sugar metabolism. The remaining EST–SSRs when searching for a putative function resulted in "no hit" (7.6%) or "hypothetical protein" (10.3%), and these may represent the specific transcriptome of sugarcane, which is yet to be characterized for its putative functions. Information of their GenBank accession numbers as well as putative functions is presented in Table 2.

To have an overview of the functional categories that those genes are distributed in, available gene ontology (GO) (The Gene Ontology Consortium 2000) annotations for best-ranked BLASTX homologies were recorded (Table 2; Fig. 1). GO presents a hierarchy of terms that vary in levels of detail and are divided among the broadest functional groups of "molecular function", "biological process", and "cellular component" categories. In the current work, we grouped the deepest level GO annotations in the terms that were two levels below the main branches of molecular function and biological process. Ninety-four clusters, 42% of 224 clusters that originated the EST–SSRs, were assigned to biological process categories, while 134 clusters, 60%, were assigned to molecular function categories. As one individual cluster can be assigned to different annotation terms, we noted that the average number of terms per cluster is 2.6 for biological process categories and 1.2 for molecular function categories (Fig. 1). It is worth noting that, as only a fraction of putative polymorphic sugarcane EST–SSRs were evaluated in the present work, it is not expected that the presented categories represent, or are proportional to, what one would obtain if the entire sugarcane genome were evaluated for the same functional categories.

Nevertheless, among the biological process categories, we could find a larger number of clusters related to macromolecule metabolic process or cellular metabolic process (Fig. 1*a*). The macromolecule metabolic process category, with 57 clusters assigned to it, represents subcategories of biopolymers, carbohydrates, and protein metabolic proc-

Fig. 2. Box plots of the distribution of the coefficients of variation of genetic similarities among all genotypes estimated by bootstrap analysis for subsamples with different numbers of EST–SSR and AFLP markers.



esses. Besides, the cellular metabolic process category, with 66 clusters assigned to it, comprises a broad spectrum of processes related to metabolic chemical pathways such as alcohol, amino acid, and cofactor metabolic processes. To accomplish their function inside a specific biological process, gene products must develop different molecular functions, as we can see in Fig. 1*b*. Eighty-six clusters already with biological process annotations were also assigned to molecular function categories (91% of 94 clusters with biological process annotations). Another 48 clusters have only molecular function annotations. We observed that the largest category was nucleic acid binding, with 43 clusters assigned to it, in which we could find 12 clusters annotated with GO terms for transcription factor activity. Another large category was associated with transferase activity, with 42 clusters assigned to, which accounts for enzymes that transfer functional molecular groups such as acyl, glycosyl, and one-carbon groups.

Polymorphism revealed by EST–SSR

Three hundred and forty-two primer pairs flanking di-, tri-, or tetrameric regions of SSRs were assayed to detect polymorphisms across the set of 18 sugarcane varieties. In 241 (70%) sequences, fragments could be amplified during PCR, producing clear and strong amplification product, while for 101 (30%) primer pairs, PCR completely failed, leading to weak amplifications or to unexpected fragment sizes. In terms of the 241 EST–SSRs, only 17 (7%) were monomorphic in the entire set of the varieties examined, whereas the other 224 were able to produce scorable polymorphisms. Out of the 17 markers that were monomorphic, nine were trinucleotide and eight were tetranucleotide motifs. In relation to the 224 polymorphic EST–SSRs, 37 (16.5%) were dinucleotides, 90 (40.2%) were trinucleotides, and 97 (43.3%) were tetranucleotides. These EST–SSRs detected 1693 alleles, ranging from 2 (ESTA19, ESTB55, ESTB91, ESTB161, ESTB163, ESTC74, ESTC75, ESTC98, ESTC100, ESTC107, and ESTC126) to 24 (ESTC119) with a mean of 7.55 alleles per EST–SSR (Table 2).

PIC values ranged from 0.16 (ESTC75) to 0.94 (ESTC119) with an average of 0.73. DP was high for the majority of EST–SSRs with an average value of 0.83, reaching the maximum value for 14 EST–SSRs (ESTA48, ESTA52, ESTA61, ESTA70, ESTB14, ESTB73, ESTB130, ESTB145, ESTB149, ESTB150, ESTC17, ESTC45, ESTC83, and ESTC129) (Table 2).

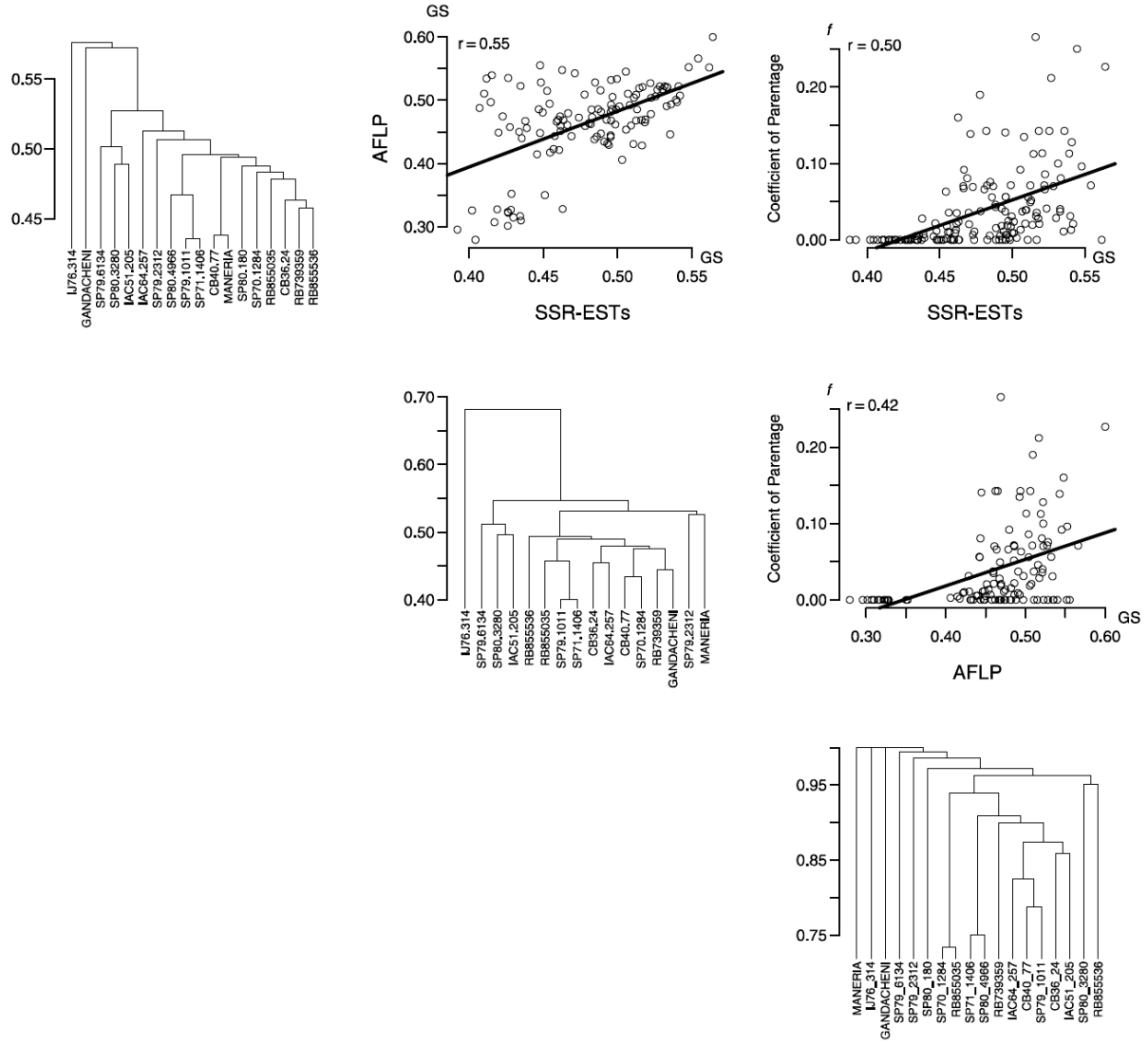
Genetic similarity using EST–SSR

Allelic data from the 224 EST–SSRs were also used to test their potential in genetic studies by ascertaining the genetic similarity in the presented sugarcane varieties. The genetic similarities based on the Jaccard coefficient were calculated with 1456 polymorphic markers amplified by the EST–SSRs (Table 3) and varied from 0.39 (between IJ76-314 (*S. officinarum*) and SP80-4966 and between IJ76-314 (*S. officinarum*) and RB855035) to 0.56 (between SP79-1011 and SP711406 and between CB40-77 and Maneria (*S. sinense*)) with a mean of 0.48 in the 153 combinations obtained. Among the species evaluated, *S. barberi* (Gandacheni) and *S. sinense* presented a similar relationship to *S. officinarum* (0.40 and 0.42, respectively) and were also related (0.45). This result is somehow in agreement with the origin of these genotypes, since *S. sinense* and *S. barberi* are usually considered natural hybrids between *Saccharum spontaneum* and *S. officinarum* (Price 1968). A genetic similarity value of 0.46 was found between SP80-180 and SP80-4966, which are parents of the Centro de Tecnologia Canavieira mapping population that was used in a genetic mapping program (Garcia et al. 2006; Oliveira et al. 2007). These results are also in agreement with those based on pedigree data, since these elite clones have some ancestors in common in the first generations, for instance, POJ2364 (Kassoer × POJ100).

Based on bootstrap analysis, it was observed that for a 5% mean coefficient of variation, which is usually considered a good value for an appropriate genetic similarity estimate, around 670 polymorphic EST–SSR markers (Fig. 2) were necessary. Using all 1456 markers characterized in this study, the mean coefficient of variation was 3.38%, which should be considered a reliable value for an appropriate genetic similarity estimate. AFLP data from Lima et al. (2002) showed that about 800 polymorphic markers were necessary to achieve a coefficient of variation of 5% (Fig. 2), indicating that fewer EST–SSR markers are necessary to obtain accurate genetic similarity estimates.

The dendrogram (Fig. 3) showed a high cophenetic coef-

Fig. 3. Diagonal: dendrograms of the genotypes constructed using the UPGMA clustering algorithm based on the genetic similarity (GS) using EST–SSR and AFLP markers and coefficient of parentage (*f*) using pedigree information. Off diagonal: dispersion plot showing the correlation (*r*) between EST–SSR-based GS, ALFP-based GS, and *f*.



ficient value (r_{cof} = 0,84), which indicates a good fit of the dendrogram with the genetic similarity matrix (Table 3). In general, the formation of groups was not very clear. However, the SP80-4966, SP79-1011, and SP71-1406 genotypes formed a group, which was in agreement with their coefficient of parentage mean (f = 0.20) when compared with the f mean of a combination of all genotypes (f = 0.04). The CB36-24, RB739359, and RB855536 genotypes were also grouped but had a mean f of 0.04. Thus, using the pedigree information, this group will not be formed. Lima et al. (2002) showed that f is not always a reliable parameter for evaluating genetic distances owing to errors in the genealogies and also because of violation of assumptions. The same occurs with CB40-77 and MANERIA genotypes that were closely grouped using EST–SSR. The correlation between AFLP-GS and SSR-GS (r = 0.55) and between SSR-GS and f (r = 0.50) was moderate. The correlation between AFLP-GS and f was moderate to low (r = 0.42), as shown by Lima et al. (2002) (Fig. 3).

Discussion

A very large number of sugarcane ESTs that are available in the SUCEST database is an important resource for the discovery of SSRs from the functional portion of the genome. In the present study, the authors have exploited this EST database that had been functionally classified based on sequence analysis to generate a set of functionally associated

markers for sugarcane. Thus, expressed sequences from sugarcane were used to develop EST–SSR markers that are expected to improve detection of marker–trait associations, since they are part of the transcribed domain of the genome. In fact, in recent years, emphasis has been shifting towards development of functional molecular markers instead of anonymous markers owing to having, for instance, a potential for assaying the functional diversity in the collection of germplasm and that may prove more useful for marker-assisted selection if found to be associated with a gene of interest.

In sugarcane, to the best of our knowledge, to date, only ~65 EST–SSR markers have been developed and described in the literature (Cordeiro et al. 2001; Pinto et al. 2004), warranting continuous efforts to develop additional new efficient functional markers for desired utility in breeding programs. In this context, the set of 224 EST–SSR markers identified in this study is expected to be a significant addition to the presently available relatively small repertoire of sugarcane functional markers. The EST–SSRs developed and characterized in the present study actually represent a subset of 2005 sugarcane EST sequences containing SSRs available in the SUCEST database. Therefore, the number of SSR markers identified and tested here is only a small proportion of what is potentially available in the SUCEST database.

A 60%–80% amplification success rate for EST–SSRs has been reported in different studies (60%, Cordeiro et al. 2001; 64%, Thiel et al. 2003; 70%, Nicot et al. 2004; 71%, Barrett et al. 2004; 80.9%, Aggarwal et al. 2007). In relation to the EST–SSRs that were herewith evaluated, 70% produced PCR products after a single round of evaluation and proved to be functional. The remaining primer pairs either failed or resulted in weak amplification. A possible explanation for this could be that one or both primers extend across a splice site or the primer pair encompasses large introns in the genomic sequence. A nonredundant EST database was used for the identification and development of the EST–SSR markers, discarding the possibility of questionable sequence information. The sequences of low quality were rejected for designing primer pairs.

The authors noticed that some differences were observed between the foreseen and the observed size of the amplification product determined from EST sequences analysis. When there was an increase in the product size, this was probably due to amplification of a small intron. If the intron was too long, the amplification product would either be out of detection range or even not be detected. As mentioned by Nicot et al. (2004), for the EST–SSRs that had a smaller amplified product than expected, this may have resulted from a small deletion within the sequence framed by the two primers, nonspecific annealing of the primers because of the low complexity of the flanking sequences, duplication of the EST sequence in the genome on the same or on another chromosome, the EST belonging to a multigenic family, and the primer being designed in the conserved domain. In fact, compared with genomic SSRs, amplicon size more frequently deviated from expectation (Cordeiro et al. 2001; Kota et al. 2001; Thiel et al. 2003; Nicot et al. 2004; Yu et al. 2004).

The EST–SSR markers developed were carried out to assess their ability to detect polymorphism among 18 sugarcane varieties. Hence, in relation to polymorphism, 65.5% of the primer pairs were polymorphic for at least one of the 18 varieties tested. The amplifications of the EST–SSRs revealed high levels of polymorphism, as indicated by the average number of alleles and the high PIC values, both typical of SSR markers, as reported by Cordeiro et al. (2000) for gSSRs in sugarcane. This suggests that most of these markers will be useful in parentage and clonal assessments because of their high potential for discrimination and also in the construction of genetic linkage maps, as these markers will likely be segregating in a range of crosses.

In general, based on visual inspection, the three dendrograms were different. However, EST–SSR and AFLP dendrograms were more alike than what was drawn using *f*. This can be verified by the coefficient correlation (*r*) between the genetic similarities, which was greater between SSR-GS and AFLP-GS (Fig. 3). Thus, different types of markers can lead to different conclusions, particularly when the number of genotypes studied is small, as described, for instance, in olive by Belaj et al. (2003) and in soybean by Powell et al. (1996*b*). According to Lima et al. (2002), the low to moderate correlation between *f* and the GS is probably due to several factors: unequal parental contribution, gene concentration in each generation of crossing, the male parent in the polycrosses being the same or a very close cultivar, or else the mistaken anotation of the parent during the generation of one of the two estimates.

The correlation between the groups formed using genetic similarities and the prediction of their performance in crosses is usually low and can vary widely from one trait to another and from one data set to another (Barbosa et al. 2003; Teklewold and Becker 2006; Guimarães et al. 2007; Qian et al. 2007). One possible explanation is the fact that random genomic DNA markers have been used for the study of genetic diversity (Lu et al. 1994; Jannoo et al. 1999*a*; Lima et al. 2002) and most of these markers may assay polymorphisms located in the noncoding regions of the genome that are poorly conserved among species. Yu et al. (2005) reported that the association between genetic similarity and hybrid performance is often inconstant and emphasized the need for specific strategies to identify parental lines with a high level of combining ability. Functional markers assay polymorphism that is associated to the coding regions of the genome and so should detect genetic diversity available inside or in regions adjacent to the genes (Thiel et al. 2003). Thus, one should expect that with the usage of EST–SSR, the correlations between genetic similarity and hybrid performance increase. More accurate predictions could be achieved using QTL mapping, which provides information about locations and genetic effects of coding regions.

EST–SSR primers have been reported to be less polymorphic compared with genomic SSRs in crop plants because of greater DNA sequence conservation in transcribed regions (Cho et al. 2000; Scott et al. 2000; Eujayl et al. 2001; Rungis et al. 2004; Russell et al. 2004; Chabane et al. 2005; Pinto et al. 2006). The rate obtained here for sugarcane (65.5%) is lower than that observed in a recent study on identification and genome mapping of EST–SSRs in kiwifruit (*Actinidia* spp.), where 93.5% of the markers were pol-

ymorphic and segregating in a mapping population (Fraser et al. 2004). Nevertheless, it was similar to the rate of EST–SSR polymorphism (66%) reported for tall fescue (Saha et al. 2004) and higher than that for wheat (53%, Nicot et al. 2004).

In the sugarcane EST–SSR database reported here, a putative function was deduced for 82.1% of the developed markers and it is expected that this number will increase in the future as the SUCEST database continuously grows. Putative functions for a significant proportion of EST–SSR markers have also been reported (Thiel et al. 2003; Gao et al. 2004; Han et al. 2004; Yu et al. 2004). The fact that ESTs containing SSRs exhibited sequence similarity to genes with a wide range of functions suggests that there is potential to identify EST–SSRs that may be directly involved in determining agronomically important characters. For example, in our results, there are ESTs with strong homology to protein involved in sugarcane metabolism. These results demonstrate that EST databases may be a valuable new source of such markers.

These markers developed from known-function genes can complement other PCR-based markers that often target noncoding regions by revealing the location and organization of gene-rich regions of the genome (Chee et al. 2004). Further, maps developed from known-function genes may be particularly valuable in species that have a large genome and (or) low polymorphism because they facilitate the use of the candidate gene mapping approach, thereby facilitating the dissection of complex traits (Pflieger et al. 2001).

The necessity of a high number of markers for the complete analysis of the sugarcane genome is justified by the complex genome of this species as well as its polyploid structure and relatively narrow genetic base, if cultivars are taken into consideration. The abundance of microsatellites in transcribed regions of the genome and the high level of polymorphism of these markers make EST libraries a valuable resource for the supply of markers for sugarcane genetic studies. These EST-derived markers characterized herewith make a substantial contribution to the understanding of the structure and function of the sugarcane genome, which can lead to improvements in sugarcane production and quality. In the long run, the development of functional and allele-specific markers for the genes controlling agronomic traits will be important for studies of genetics and breeding of sugarcane.

Acknowledgements

The authors gratefully acknowledge Ms. Itaraju Junior Baracuhy Brum for his constructive and crucial assistance with functional analysis. This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (02/01167-1) and from CTC (Centro de Tecnologia Canavieira – Piracicaba/SP). K.M.O. received a doctorate fellowship from FAPESP (02/00197-4) and also received a PDEE fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (0181-05-6). L.R.P. received a postdoctorate fellowship from FAPESP (01/14656), T.G.M. received an IC fellowship from FAPESP (04/10596-9), and A.P.S. and A.A.F.G. received research fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Aggarwal, R.K., Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V., and Singh, L. 2007. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor. Appl. Genet. **114**: 359–372. doi:10.1007/s00122-006-0440-x. PMID:17115127.

Atkinson, B., and Therneau, T. 2007. Kinship: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees, R package version 1.1.0-17. Available from http://cran.r-project.org/web/packages/kinship/index.html.

Barbosa, A.M.M., Geraldi, I.O., Benchimol, L.L., Garcia, A.A.F., Souza, C.L.S., Jr., and Souza, A.P. 2003. Relationship of intra- and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. Euphytica, **130**: 87–99. doi:10.1023/A:1022381008721.

Barrett, B., Griffiths, A., Schreiber, M., Ellison, N., Mercer, C., Bouton, J., et al. 2004. A microsatellite map of white clover. Theor. Appl. Genet. **109**: 596–608. PMID:15103407.

Bhat, P.R., Krishnakumar, V., Hendre, P.S., Rajendrakumar, P., Varshney, R.K., and Aggarwal, R.K. 2005. Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interespecific hybrid of *Coffee canephora* × *Coffee congensis*). Mol. Ecol. Notes, **5**: 80–83. doi:10.1111/j.1471-8286.2004.00839.x.

Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L., and Trujillo, I. 2003. Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theor. Appl. Genet. **107**: 736–744. doi:10.1007/s00122-003-1301-5. PMID:12819908.

Carson, D.L., and Botha, F.C. 2000. Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci. **40**: 1769–1779.

Casu, R., Dimmock, C., Thomas, M., Bower, N., Knight, D., Grof, C., et al. 2001. Genetic and expression profiling in sugarcane. Proc. Int. Soc. Sugarcane Technol. **24**: 626–627.

Casu, R., Manners, J., Bonnett, G., Jackson, P., McIntyre, C., Dunne, R., et al. 2005. Genomics approaches for the identification of genes determining important traits in sugarcane. Field Crops Res. **92**: 137–147. doi:10.1016/j.fcr.2005.01.029.

Chabane, K., Ablett, G.A., Cordeiro, G.M., Valkoun, J., and Henry, R.J. 2005. EST versus genomics derived microsatellite markers for genotyping wild and cultivated barley. Genet. Resour. Crop Evol. **52**: 903–909. doi:10.1007/s10722-003-6112-7.

Chee, P.W., Rong, J., Williams-Coplin, D., Schulze, S.R., and Paterson, A.H. 2004. EST derived PCR-based markers for functional gene homologues in cotton. Genome, **47**: 449–462. doi:10.1139/g04-002. PMID:15190362.

Chen, C., Zhou, P., Choi, Y.A., Huang, S., and Jr-Gmitter, F.G. 2006. Mining and characterizing microsatellites from citrus ESTs. Theor. Appl. Genet. **112**: 1248–1257. doi:10.1007/s00122-006-0226-1. PMID:16474971.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., et al. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100**: 713–722. doi:10.1007/s001220051343.

Cordeiro, G.M., Taylor, G.O., and Henry, R.J. 2000. Characterization of microsatellite markers from sugarcane (*Saccharum* spp.), a highly polyploid species. Plant Sci. **155**: 161–168. doi:10.1016/S0168-9452(00)00208-9. PMID:10814819.

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M., and Henry, R.J. 2001. Microsatellite markers from sugarcane (*Sac-*

charum spp) ESTs cross-transferable to *Erianthus* and *Sorghum*. Plant Sci. **160**: 1115–1123. doi:10.1016/S0168-9452(01)00365-X. PMID:11337068.

Cox, T.S., Kiang, Y.T., Gorman, M.B., and Rogers, D.M. 1985. Relationships between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci. **25**: 529–532.

Creste, S., Tulmann-Neto, A., and Figueira, A. 2001. Detection of single repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol. Biol. Rep. **19**: 299–306. doi:10.1007/BF02772828.

Edwards, K.J., Barker, J.H., Daly, A., Jones, C., and Karp, A. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques, **20**: 758–760. PMID:8723911.

Efron, B., and Tibshirani, R.J. 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Eujayl, I., Sorrells, M., Baum, M., Wolters, P., and Powell, W. 2001. Assessment of genotypic variation among cultivated durum wheat based on EST–SSRs and genomic SSRs. Euphytica, **119**: 39–43. doi:10.1023/A:1017537720475.

Fraser, L.G., Harvey, C.F., Crowhurst, R.N., and De Silva, H.N. 2004. EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theor. Appl. Genet. **108**: 1010–1016. doi:10.1007/s00122-003-1517-4. PMID:15067386.

Gao, L.F., Jing, R.L., Huo, N.X., Li, X.P., Zhou, R.H., Chang, X.P., et al. 2004. One hundred and one new microsatellite loci derived from ESTs (EST–SSRs) in bread wheat. Theor. Appl. Genet. **108**: 1392–1400. doi:10.1007/s00122-003-1554-z. PMID:14968301.

Garcia, A.A.F., Kido, E.A., Meza, A.N. Souza, H.M.B., Pinto, L.R., Pastina, M.M., Leite, C.S., da Silva, J.A.G., Ulian, E.C., Figueira, A., and Souza, A.P. 2006. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor. Appl. Genet. **112**: 298–314. doi:10.1007/s00122-005-0129-6. PMID:16307229.

Graham, J., Smith, K., MacKenzie, K., Jorgenson, L., Hackett, C., and Powell, W. 2004. The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp. *idaeus*) based on AFLPs, genomic-SSR and EST–SSR markers. Theor. Appl. Genet. **109**: 740–749. doi:10.1007/s00122-004-1687-8. PMID:15127204.

Guimarães, P.S., Paterniani, M.E.A.G.Z., Lüders, R.R., Souza, A.P., Laborda, P.R., and Oliveira, K.M. 2007. Correlação da heterose de híbridos de milho com divergência genética entre linhagens. Pesq. Agropec. Bras. **42**: 811–816.

Gupta, P.K., and Varshney, R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, **113**: 163–185. doi:10.1023/A:1003910819967.

Gupta, P.K., Rutsgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H.S. 2003. Transferable EST–SSR markers for study of polymorphism and genetic diversity in bread wheat. Mol. Genet. Genomics, **270**: 315–323. doi:10.1007/s00438-003-0921-4. PMID:14508680.

Han, Z.G., Guo, W.Z., Song, X.L., and Zhang, T.Z. 2004. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol. Genet. Genomics, **272**: 308–327. doi:10.1007/s00438-004-1059-8. PMID:15368122.

Han, Z.G., Wang, C., Song, X., Guo, W., Gou, J., Li, C., et al. 2006. Characteristics, development and mapping of *Grossypium hirsutum* derived EST–SSRs in allotetraploid cotton. Theor. Appl. Genet. **112**: 430–439. doi:10.1007/s00122-005-0142-9. PMID:16341684.

Hoisington, D., Khairallah, M., and González-de-Léon, D. 1994. Laboratory protocols: CIMMYT Applied Molecular Genetics Laboratory. CYMMYT, México.

Ihaka, R., and Gentleman, R. 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. **5**: 299–314. doi:10.2307/1390807.

Jaccard, P. 1901. Étude comparative de la distribuition florale dans une portion des Alpes et des Jura. Bull. Soc. Vandoise Sci. Nat. **37**: 547–579.

Jannoo, N., Grivet, L., Seguin, M., Paulet, F., Domaingue, R., Rao, P.S., et al. 1999*a*. Molecular investigation of the genetic base of sugarcane cultivars. Theor. Appl. Genet. **99**: 171–184. doi:10.1007/s001220051222.

Kantety, R.V., Rota, M.L., Matthews, D.E., and Sorrells, M.E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. **48**: 501–510. doi:10.1023/A:1014875206165. PMID:11999831.

Kempthorne, O. 1957. An introduction to genetic statistics. John Wiley & Sons, New York.

Khlestkina, E.K., Than, M.H.M., Pestsova, E.G., Röder, M.S., Malyshev, S.V., Korzu, V., and Börner, A. 2004. Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor. Appl. Genet. **109**: 725–732. doi:10.1007/s00122-004-1659-z. PMID:15300380.

King, G., Nienhuis, J., and Hussey, C. 1993. Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theor. Appl. Genet. **86**: 1028–1032.

Kota, R., Varshney, R.K., Thiel, T., Dehmer, K.J., and Graner, A. 2001. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas, **135**: 145–151. doi:10.1111/j.1601-5223.2001.00145.x. PMID:12152327.

Lima, M.L.A., Garcia, A.A.F., Oliveira, K.M., Matsuoka, S., Souza, C.L., Jr., and Souza, A.P. 2002. Analysis of genetics similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp.). Theor. Appl. Genet. **104**: 30–38. doi:10.1007/s001220200003. PMID:12579425.

Lu, Y.H., D'Hont, A., Paulet, F., Grivet, L., Arnaud, M., and Glaszmann, J.C. 1994. Molecular diversity and genome structure in modern sugarcane varieties. Euphytica, **78**: 217–226. doi:10.1007/BF00027520.

Mateescu, R.G., Zhang, Z., Tsai, K., Phavaphutanon, J., Burton-Wurster, N.I., Lust, G., et al. 2005. Analysis of allele fidelity, polymorphic information content, and density of microsatellites in a genome-wide screening for hip dysplasia in a crossbreed pedigree. J. Hered. **96**: 847–853. doi:10.1093/jhered/esi109. PMID:16251522.

Nicot, N., Chiquet, V., Gandon, B., Amilhat, L., Legeai, F., Leroy, P., et al. 2004. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor. Appl. Genet. **109**: 800–805. doi:10.1007/s00122-004-1685-x. PMID:15146317.

Oliveira, K.M., Pinto, L.R., Marconi, T.G., Margarido, G.R.A., Pastina, M.M., Teixeira, L.H., et al. 2007. Functional genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. Mol. Breed. **20**: 189–208. doi:10.1007/s11032-007-9082-1.

Pashley, C.H., Ellis, J.R., McCauley, D.E., and Burke, J.M. 2006. EST database as a source for molecular markers: lessons from *Helianthus*. J. Hered. **97**: 381–388. doi:10.1093/jhered/esl013. PMID:16840524.

Pflieger, S., Lefebvre, V., and Causse, M. 2001. The candidate gene approach in plant genetics: a review. Mol. Breed. **7**: 275–291. doi:10.1023/A:1011605013259.

Published by NRC Research Press

Pillen, K., Binder, A., Kreuzkam, B., Ramsay, L., Waugh, R., Förster, J., and Leon, J. 2000. Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theor. Appl. Genet. **101**: 652–660. doi:10.1007/s001220051527.

Pinto, L.R., Oliveira, K.M., Ulian, E.C., Garcia, A.A.F., and Souza, A.P. 2004. Survey in the expressed sequence tag database (SUCEST) for simple sequence repeats. Genome, **47**: 795–804. doi:10.1139/g04-055. PMID:15499394.

Pinto, L.R., Oliveira, K.M., Marconi, T.G., Garcia, A.A.F., Ulian, E.C., and Souza, A.P. 2006. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed. **125**: 378–384. doi:10.1111/j.1439-0523.2006.01227.x.

Powell, W., Machray, G.C., and Provan, J. 1996*a*. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. **1**: 215–221.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalsky, A. 1996*b*. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. **2**: 225–238. doi:10.1007/BF00564200.

Price, S. 1968. Cytology of Chinese and North Indian sugarcanes. Econ. Bot. **22**: 155–164.

Qian, W., Sass, O., Meng, J., Li, M., Frauen, M., and Jung, C. 2007. Heterotic patterns in rapeseed (*Brassica napus* L.): I. Crosses between spring and Chinese semi-winter lines. Theor. Appl. Genet. **115**: 27–34. doi:10.1007/s00122-007-0537-x. PMID:17453172.

Raboin, L.-M., Oliveira, K.M., Lecunff, L., Telismart, H., Roques, D., Butterfield, M., et al. 2006. Genetic mapping in the high polyploid sugarcane using a bi-parental progeny: identification of a gene controlling stalk colour and new rust resistance gene. Theor. Appl. Genet. **112**: 1382–1391. doi:10.1007/s00122-006-0240-3. PMID:16552554.

Rossetto, M., McNally, J., and Henry, R.J. 2002. Evaluating the potential of SSR: flanking regions examining taxonomic relationships in the Vitaceae. Theor. Appl. Genet. **104**: 61–66. doi:10.1007/s001220200007. PMID:12579429.

Rudd, S. 2003. Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. **8**: 321–329. doi:10.1016/S1360-1385(03)00131-6. PMID:12878016.

Rungis, D., Bérubé, Y., Zhang, J., Ralph, S., Ritland, C.E., Ellis, B.E., et al. 2004. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor. Appl. Genet. **109**: 1283–1294. doi:10.1007/s00122-004-1742-5. PMID:15351929.

Russell, J., Booth, A., Fuller, J., Harrower, B., Hedley, P., Machray, G., and Powell, W. 2004. A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley. Genome, **47**: 389–398. PMID:15060592.

Saha, M.C., Mian, M.A.R., Eujayl, I., Zwonitzer, J.C., Wang, L.J., and May, G.D. 2004. Tall fescue EST–SSR markers with transferability across several grass species. Theor. Appl. Genet. **109**: 783–791. doi:10.1007/s00122-004-1681-1. PMID:15205737.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723–726. doi:10.1007/s001220051344.

Sharopova, N., McMullen, M.D., Schultz, L., Schroeder, S., Sanchez-Villeda, H., Gardiner, J., et al. 2002. Development and mapping of SSR markers for maize. Plant Mol. Biol. **48**: 463–481. doi:10.1023/A:1014868625533. PMID:12004892.

Sneath, P.H.A., and Sokal, R.R. 1973. Numerical taxonomy. Freeman, San Francisco, Calif.

Teklewold, A., and Becker, H.C. 2006. Comparison of phenotypic and molecular distance to predict heterosis and F1 performance in Ethiopian mustard (*Brassica carinata* A. Braum). Theor. Appl. Genet. **112**: 752–759. doi:10.1007/s00122-005-0180-3. PMID:16365759.

Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., et al. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100**: 697–712. doi:10.1007/s001220051342.

Tessier, C., David, J., This, P., Boursiquot, J.M., and Charrier, A. 1999. Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor. Appl. Genet. **98**: 171–177. doi:10.1007/s001220051054.

The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. Nat.e Genet. **25**: 25–29.

Thiel, T., Michalek, W., Varshney, R.K., and Graner, A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **106**: 411–422. PMID:12589540.

Tivang, J.G., Nienhuis, J., and Smith, O.S. 1994. Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor. Appl. Genet. **89**: 259–264. doi:10.1007/BF00225151.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and Graner, A. 2002. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell. Mol. Biol. Lett. **7**: 537–546. PMID:12378259.

Varshney, R.K., Graner, A., and Sorrells, M.E. 2005*a*. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. **23**: 48–55. doi:10.1016/j.tibtech.2004.11.005. PMID:15629858.

Varshney, R.K., Sigmund, R., Borner, A., Korzun, V., Stein, N., Sorrells, M.E., and Graner, A. 2005*b*. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci. **168**: 195–202. doi:10.1016/j.plantsci.2004.08.001.

Warnke, S.E., Barker, R.E., Jung, G., Sim, S., Mian, M.A.R., Saha, M.C., et al. 2004. Genetic linkage mapping of an annual × perennial ryegrass population. Theor. Appl. Genet. **109**: 294–304. doi:10.1007/s00122-004-1647-3. PMID:15071730.

Yu, C.Y., Hu, S.W., Zhao, H.X., and Guo, A.G. 2005. Genetic distances revealed by morphological characters, isozymes, protein and RAPD markers and their relationships with hybrid performance in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. **110**: 511–518. doi:10.1007/s00122-004-1858-7. PMID:15578151.

Yu, J.K., Dake, T.M., Singh, S., Benscher, D., Li, W., Gill, B., and Sorrells, E. 2004. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome, **47**: 805–818. doi:10.1139/g04-057. PMID:15499395.

ARTIGO IV

"Functional markers for gene mapping and genetic diversity studies in sugarcane"

T.G. Marconi, E.A. Costa, H.R.C.A.N. Miranda, M.C. Mancini, C.B. Cardoso-Silva, K.M. Oliveira, L.R. Pinto,M. Mollinari, A.A.F. Garcia, A.P. Souza

Publicado na revista

BMC Research notes (4:264; 2011)

Marconi *et al. BMC Research Notes* 2011, **4**:264
http://www.biomedcentral.com/1756-0500/4/264

BMC Research Notes

SHORT REPORT Open Access

Functional markers for gene mapping and genetic diversity studies in sugarcane

Thiago G Marconi[1], Estela A Costa[1], Hercília RCAN Miranda[1], Melina C Mancini[1], Cláudio B Cardoso-Silva[1], Karine M Oliveira[1,5], Luciana R Pinto[2], Marcelo Mollinari[3], Antônio AF Garcia[3] and Anete P Souza[1,4*]

Abstract

Background: The database of sugarcane expressed sequence tags (EST) offers a great opportunity for developing molecular markers that are directly associated with important agronomic traits. The development of new EST-SSR markers represents an important tool for genetic analysis. In sugarcane breeding programs, functional markers can be used to accelerate the process and select important agronomic traits, especially in the mapping of quantitative traits loci (QTL) and plant resistant pathogens or qualitative resistance loci (QRL). The aim of this work was to develop new simple sequence repeat (SSR) markers in sugarcane using the sugarcane expressed sequence tag (SUCEST database).

Findings: A total of 365 EST-SSR molecular markers with trinucleotide motifs were developed and evaluated in a collection of 18 genotypes of sugarcane (15 varieties and 3 species). In total, 287 of the EST-SSRs markers amplified fragments of the expected size and were polymorphic in the analyzed sugarcane varieties. The number of alleles ranged from 2-18, with an average of 6 alleles per locus, while polymorphism information content values ranged from 0.21-0.92, with an average of 0.69. The discrimination power was high for the majority of the EST-SSRs, with an average value of 0.80. Among the markers characterized in this study some have particular interest, those that are related to bacterial defense responses, generation of precursor metabolites and energy and those involved in carbohydrate metabolic process.

Conclusions: These EST-SSR markers presented in this work can be efficiently used for genetic mapping studies of segregating sugarcane populations. The high Polymorphism Information Content (PIC) and Discriminant Power (DP) presented facilitate the QTL identification and marker-assisted selection due the association with functional regions of the genome became an important tool for the sugarcane breeding program.

Background

Sugarcane (*Saccharum* spp.) is a member of a polyploid complex belonging to the Andropogoneae tribe in the Poaceae family. This crop is economically important because it is the main source of both sugar and alcohol, and it accounts for two thirds of the world's sugar production [1]. In Brazil, the production about 625 million tons of sugarcane is estimated for the 2010/2011 harvest, surpassing the production of the last harvest [2].

Modern sugarcane cultivars are highly polyploid and often aneuploid, with chromosome numbers ranging from 100 to 130 [3]. These modern cultivars are interspecific hybrids derived essentially from crosses between *S. officinarum* (2n = 80, x = 10), a species that has stalks with high sugar content, and *S. spontaneum* (2n = 40-128, x = 8), a wild and vigorous species that is resistant to several sugarcane diseases. Because of sugarcane's interspecific origins and its high polyploid chromosome number, crosses between different sugarcane varieties produce aneuploid progeny [4,5]. These characteristics make the time required for developing new varieties as high as 12 to 15 years [6], and they represent some of the main drawbacks in sugarcane breeding programs.

Molecular markers associated with important agronomic traits can be used to help select varieties at early stages of breeding programs, known as Marker-Assisted Selection (MAS) and to choose the best parents in a

* Correspondence: anete@unicamp.br
[1]Centro de Biologia Molecular e Engenharia Genética (CBMEG) - Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, CEP 13083-970, Campinas, SP, Brazil
Full list of author information is available at the end of the article

Marconi *et al. BMC Research Notes* 2011, **4**:264
http://www.biomedcentral.com/1756-0500/4/264

cross. Therefore, they could facilitate a significant reduction in the time and cost involved to develop new varieties, and they should help to bypass barriers in sugarcane breeding.

Simple sequence repeats (SSR) or microsatellites are short fragments of DNA that consist of small motifs of one to six tandem repeat base pairs that are flanked by well-conserved sequences, and they allow for the design of specific primers [7-9]. The SSRs developed from expressed sequence tags (ESTs) are derived from expressed genes. The advantage of using ESTs is that they can facilitate the mapping of genes with known functions pathways [10]. They are also located in the transcribed portion of the genome [11], which allows for a direct association between genes and important agronomic traits. ESTs have been developed for a large number of plant species, including cotton [12], wheat [13], barley [14], and the rubber tree [15].

To identify available ESTs, our group used the Brazilian EST database (Sugarcane EST Project- SUCEST), which is the largest database of sugarcane ESTs; it contains 237,954 ESTs grouped in 43,000 clusters [16]. Most of these clusters represent genes associated with important metabolic processes such as photosynthesis, carbohydrate metabolism, sugar transport, amino acid metabolism, and biotic and abiotic stress response mechanisms.

A preliminary analysis involving the development of functional SSRs markers from the SUCEST database was reported by Da Silva [17], who identified 20 EST-SSRs revealing this database as a good source for the development of molecular markers.

In total, our group has developed 837 sugarcane EST-SSRs. Initially were reported the development and characterization of 30 EST-SSRs [18]. Later, developed 100 more EST-SSRs and compared them with 50 other SSRs derived from genomic libraries (gSSR) [19]. After, developed 149 additional EST-SSRs [20], and more recently, developed 193 EST-SSRs [21]. In the present article, we develop and characterize 365 novel EST-SSRs.

The aim of the present work was to characterize 365 SSR sugarcane markers identified in the SUCEST database. Appropriate primers and PCR amplification conditions were developed for each SSR, and their polymorphism information content (PIC) and discrimination power (DP) were determined. To validate the newly developed EST-SSRs, the genetic diversity among 18 sugarcane genotypes (15 varieties and 3 species) was evaluated and the results were compared with previously reported data that utilized AFLPs and pedigrees [22].

Methods

Plant Material and DNA Extraction

A total of 13 commercial sugarcane varieties (*CB 36-24, SP 79-1011, CB 40-77, IAC 51-205, RB 73-9359, SP 70-1284, IAC 64-257, SP 71-1406, SP 80-3280, RB 85-5035, SP 79-6134, SP 79-2312*, and *RB 85-5536*) and two parental varieties (*SP 80-180* and *SP 80-4966*) from a genetic mapping program [20,23] at Centro de Tecnologia Canavieira (CTC - Piracicaba, SP, Brazil) and one clone from each of three *Saccharum* species (*IJ76-314, S. officinarum*; *Gandacheni, S. barberi*; and *Maneria, S. sinense*) (Table 1) were used to validate the developed EST-derived markers and to screen for polymorphic SSRs. The studied germplasm was chosen according to a previous study on sugarcane genetic diversity that utilized AFLP markers and pedigree information [22]. Genomic DNA was extracted from 300 mg of lyophilized young leaf tissue using the method described by Hoisington et al. [24] with minor modifications.

Development of SSR markers

SSR markers were designed from 2005 sugarcane consensus sequences derived from the SUCEST database (The Sugarcane EST project), as described in detail by Pinto et al. [18]. A set of 365 locus-specific primers flanking SSRs with trinucleotide motifs was designed using the Primer Select DNAStar® package software. The primers were synthesized by Integrated DNA Technologies (IDT) (São Paulo, SP, Brazil). The criteria used to design the primers for SSR flanking regions were the same as those described by Cordeiro et al. [25].

To evaluate the importance of these EST-SSR markers, they were compared to those in the GenBank database using BLASTX directly from NCBI. The biological function was identified using the Gene Ontology database.

PCR amplification and EST-SSR visualization

PCR amplifications were performed in 20 μl reaction containing 10× PCR buffer (10 mM Tris-HCl, 50 mM KCl), 2.0 mM $MgCl_2$, 100 μM of each dNTP, 0.2 μM of each forward and reverse primer, 0.5 U Taq DNA polymerase (Invitrogen, SP, Brazil), and 40 ng of template DNA. After an initial denaturation at 94°C for 3 min, DNA fragments were PCR-amplified for 31 cycles of 1 min at 94°C, 1 min at the annealing temperature specific to each primer, and 1 min at 72°C; the final extension time was for 2 min at 72°C. The annealing temperature for each primer pair was determined using a temperature gradient ranging from 50 to 65°C on a PTC-200 thermocycler (MJ Research). The amplification products were separated by electrophoresis on 6% denaturing polyacrylamide gels in 1× TBE buffer. To perform the electrophoresis, PCR reaction mixtures were mixed with equal volumes of loading buffer (formamide containing 0.8 mM EDTA and traces of bromophenol blue and xylene cyanol) and denatured at 90°C for 3 min. Samples were then loaded on a pre-heated S2001 Model (Life Technologies). A 10 bp ladder was used (Invitrogen, SP, Brazil) as a size standard. The DNA

Marconi *et al. BMC Research Notes* 2011, **4**:264
http://www.biomedcentral.com/1756-0500/4/264

Table 1 Sugarcane genotypes used for marker validation with pedigree relationships covering two generations and their origins

Genotypes	Pedigree	Origin
CB36-24	[POJ2364 × EK28] POJ2878 × ?	Campos, Brazil
CB40-77	[POJ2364 × EK28] POJ2878 × CO290 [CO221 × D74]	Campos, Brazil
SP79-1011	[CO419 × CO419] NA5679 × CO775 [POJ2878 × CO371]	São Paulo, Brazil
SP70-1284	[POJ2878 × ?] B4176 × ?	São Paulo, Brazil
SP71-1406	[CO419 × CO419] NA5679 × ?	São Paulo, Brazil
SP80-3280	[CP5530 × CP5376] SP71-1088 × H575028 [H49134 × ?]	São Paulo, Brazil
SP79-6134	[H53263 × H507209] H634644 × ?	São Paulo, Brazil
SP79-2312	[SP71-6106 × ?] CP5659 × ?	São Paulo, Brazil
SP80-4966	[NA5679 × ?] SP711406 × ?	São Paulo, Brazil
SP80-180	[B3337 × ?] B3337 × ?	São Paulo, Brazil
IAC51-205	[POJ2364 × EK28] POJ2878 × ?	Campinas, Brazil
IAC64-257	[POJ2878 × CO290] CO419 × IAC49-131 [CP27108 × ?]	Campinas, Brazil
RB739359	[CO419 × MZ336] IANE5534 × ?	Republic of Brazil
RB855035	[CP521 × CP48103] L6014 × SP701284 [CB4176 × ?]	Republic of Brazil
RB855536	[IAC48-65 × ?] SP70-1143 × RB72454 [CP5376 × ?]	Republic of Brazil
Gandacheni (*S. barberi*)	*S. barberi* × ?	Saretha, Índia
IJ76-314 (*S. officinarum*)	*S. officinarum* × ?	Iryan, Java
Maneria (*S. sinense*)	*S. sinense* × ?	Pansahi, China

fragments were visualized by silver staining according to Creste et al. [26].

Statistical analysis

Polymorphism analysis

All segregating unique bands were treated as dominant markers because sugarcane is a hybrid species that incorporates multiple chromosomes sets that can pair and recombine freely, that is, with little or no preferential pairing. The data were scored based on the presence (1) or absence (0) of a band for each of the 18 sugarcane genotypes. Both PIC and DP were obtained for a polymorphism analysis of each locus. PIC values were calculated based on the following formula:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where p_i is the frequency of the i^{th} allele and the summation extends over n alleles [27]. DP values for the k^{th} primer were calculated based on the following formula:

$$DP_k = 1 - \sum_{j=1}^{l} p_j \frac{Np_j - 1}{N - 1},$$

where N is the number of individuals and p_j is the frequency of the j^{th} pattern [28]. PIC was used as a tool to measure the information that a given marker locus could provide for the pool of genotypes, while DP was used as a quantification tool to measure the efficiency of a given marker for the discrimination of genotypes; i.e., the probability that two randomly chosen individuals have different patterns.

Genetic similarity analysis

To evaluate the genetic similarity among all the genotypes, the polymorphic bands were used to construct a binary matrix. The EST-SSR-based genetic similarity (SSR-GS) among all of the genotypes was estimated according to the Jaccard's similarity coefficient [29]. The corresponding genetic similarity matrix was used to generate a dendrogram based on the Unweighted Pair Group Method with the Arithmetic Average (UPGMA) algorithm, as suggested by Sneath and Sokal [30]. All analyses were carried out using NTSYSpc 2.11X [31]. A bootstrap analysis with 10,000 random samplings was applied to estimate the reliability of the dendrogram branches using BOOD version 3.0 [32].

Results and Discussion

Development of potentially functional SSR markers

According to the search criteria adopted, a total of 2005 clusters containing SSRs were found in the SUCEST database. These sequences were considered useful because they contained dinucleotide, trinucleotide or tetranucleotide repeats as described by literature [18]. Oligonucleotide primer pairs were successfully synthesized and tested for 365 of the ESTs that contained trinucleotide motifs. The sequences were selected based on the size of the individual repeat units and the expected homology to known genes. The EST-SSRs were initially tested in agarose gels to verify the presence and quality of the product amplification of a predicted size. Afterwards, the selected EST-SSRs were

Marconi *et al. BMC Research Notes* 2011, **4**:264
http://www.biomedcentral.com/1756-0500/4/264

screened in 6% denaturing polyacrylamide gels to identify a high-quality banding pattern. The details regarding the primer pair sequences and the product size of the SSR motifs for the selected markers are described in Additional file 1.

The used molecular markers derived from EST is a strategy that allows a direct mapping interest genes [19,20,33-35] and the discovery of the functional portion of the genome [21]. In this study, a large number of sugarcane ESTs derived from the SUCEST database were exploited to generate a set of functionally associated markers for sugarcane. Expressed sequences from sugarcane were used to develop EST-SSR markers that are expected to improve the detection of marker-trait associations because they are part of the transcribed domain of the genome. In recent years, emphasis has been placed on the development of functional molecular markers in order to assess variation in wild relatives or to establish genetic similarities between genotypes for breeding purposes, it is valuable to have information regarding the variability of specific gene that potentially affect important breeding traits.

Functional characterization EST-SSR

Of the 197 ESTs sequences compared to sequences from GenBank using the BLASTX algorithm, 65.98% had homology with a known protein and 23.62% had homology with a putative protein. The remaining 10.40% of the sequences did not possess homology with any sequence in the database.

The unigenes identified in this study had significant similarity with genes from other plant species. The majority of the similarities were found to *Sorghum bicolor*, *Oryza sativa*, and *Zea mays*. The enrichment of similarities with genomes these species is likely because they have sequenced genomes, a large number of expressed sequences deposited in databases, and a close phylogenetic relationship with sugarcane. Data regarding the homologies to each of the EST-SSR markers are shown in Additional file 2.

Ontology has been used in many model organisms to annotate the function of genes and their products and may be represented in terms of Molecular Function (MF), Biological Process (BP) and Cellular Component (CC). We made notes of the MF and BP for each EST. Based on the 130 ESTs that had homology with a known protein, 97 (74.61%) of these ESTs possessed a known molecular function and/or biological process. Of these, 78 (59.54%) ESTs possessed a known biological process and 88 (67.18%) ESTs possessed a known molecular function. The molecular functions most occurring were characterized into the following groups: binding (42.86%), transcription factor activity (11.61%) and kinase activity (8.04%). Regarding the biological processes, the ESTs characterized in this work are mainly involved in cellular metabolic process (35.63%), regulation of transcription (18.39%) and transport (17.24%). The ontology category, based on the molecular function and biological processes of EST-SSR markers, is represented graphically in Figure 1. The functional annotation (MF and BP) for each EST-SSR marker is listed in additional file 2.

We identified several EST-SSR markers that can be associated with genes that play important physiological roles in plants. Some of the clusters that gave rise to these markers shared homology with were related to proteins deposited in GenBank. For example, the SCB207 marker is derived from an EST with homology to a gene (ERECTA) that regulates the efficiency of sweating. This gene coordinates sweating and photosynthesis [36] and indirectly influences productivity. The SCB218 marker is associated with the acid soluble invertase protein, an enzyme that limits the accumulation of sugars such as sucrose in the immature internodes of sugarcane stalks [37]. We also identified SCB324 and SCB336 as markers associated with the MoS2 protein. MoS2 was identified in *Arabdopsis thaliana* and is essential in the innate immune response, especially at the beginning of the cascade that is activated in response to pathogens [38]. Finally, the SCB487 marker may be associated with a gene for sugar transport.

Polymorphism revealed by EST-SSR

A total of 365 primer pairs flanking SSR regions were designed to detect polymorphisms across the set of 18 sugarcane genotypes (15 genotypes came from a interspecific hybrid *Saccharum spp*., and the remaining from 3 different species: *S. officinarum, S. barberi*, and *S. sinense*). In 287 sequences (78.6%) of the total developed, clear, strong DNA fragments were produced by PCR amplification. The PCR reactions using the remaining 78 (21.4%) primer pairs failed, resulting in weak DNA amplification or an unexpected fragment size. Of the 287 selected primer pairs that produced a clear product, 197 (68.6%) were polymorphic and 90 (31.4%) were monomorphic. The EST-SSR amplifications revealed high levels of polymorphism and detected 1261 alleles with a range of 2 alleles (SCB168, SCB180, SCB187, SCB192, SCB193, SCB195, SCB204, SCB210, SCB231, SCB232, SCB234, SCB240, SCB241, SCB250, SCB293, SCB302, SCB306, SCB345, SCB377, SCB377, SCB379, SCB402, SCB405, SCB407, SCB416, SCB418, SCB445, SCB449 and SCB459) to 18 alleles (SCB344), the mean was 6.4 alleles per EST-SSR (Additional file 1).

The PIC values ranged from 0.21 (SCB345) to 0.92 (SCB344), and the average PIC was 0.69. The DP was high for the majority of the EST-SSRs, with an average value of 0.80. A maximum value was found for eight EST-SSRs (SCB181, SCB246, SCB285, SCB330, SCB334, SCB373, SCB374 and SCB466) (Additional file 1).

Marconi *et al. BMC Research Notes* 2011, **4**:264
http://www.biomedcentral.com/1756-0500/4/264

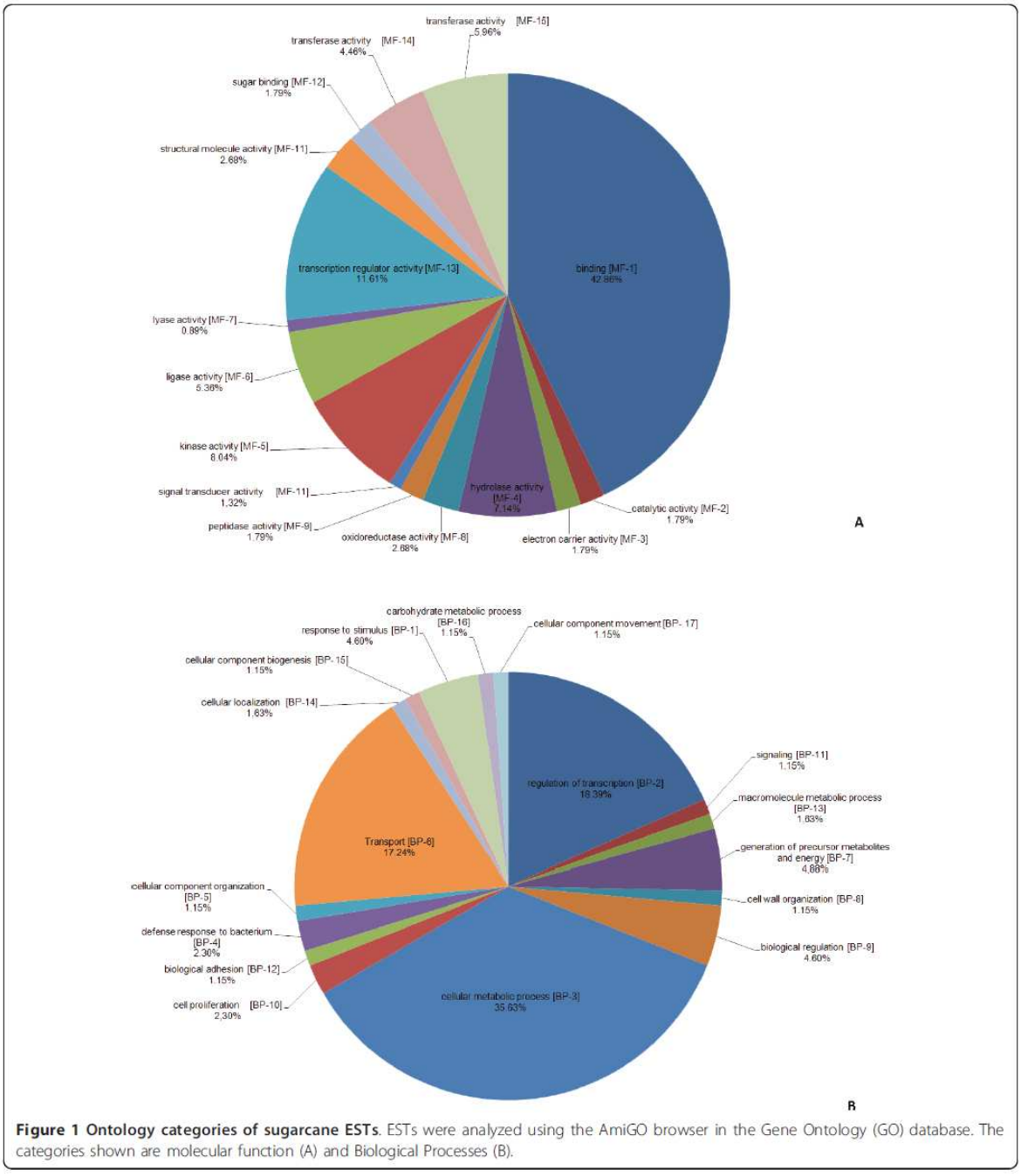


Figure 1 Ontology categories of sugarcane ESTs. ESTs were analyzed using the AmiGO browser in the Gene Ontology (GO) database. The categories shown are molecular function (A) and Biological Processes (B).

Similar PIC values were reported by Oliveira et al. [21] using the same set of sugarcane varieties ranged from 0.16 to 0.94 with an average 0.73. While Cordeiro et al. [39], considering sugarcane Hybrids, *S. officinarum* and *S. sinense*, found PIC values 0.77, 0.68 and 0.70 respectively, suggesting that most of these markers will be useful in the assessment of the genetic diversity in sugarcane because of their high potential for discrimination and genetic mapping.

The amplification success rate of the EST-SSRs was 78.6% and is similar to those rates reported in previous studies with sugarcane (60%, [25]; 77%, [18]; 62%, [19];

Marconi *et al. BMC Research Notes* 2011, **4**:264
http://www.biomedcentral.com/1756-0500/4/264

70%, [21]), barley (64%, [14]), wheat (70%, [40]), white clover (71%, [41]) and coffee (80.9%, [42]). The remaining (21.4%) primer pairs failed to amplify DNA or resulted in weak DNA amplification. The amplification of a larger fragment than expected could reflect the presence of introns within the genomic DNA sequence. The lack of amplification could reflect the presence of long introns between the sequences homologous to the primers in the genomic DNA. Other reasons for a failed amplification include the absence of the allele or the divergence of the sequence in the SSR flanking region.

The high values obtained are typical of SSR markers, and these SSRs may be capable of detecting larger variation as reported in sugarcane gSSRs [19,43] and sugarcane EST-SSRs [18,21]. Most of the markers isolated will be useful in determining sugarcane parentage and making clonal assessments; the high levels of polymorphism indicate that they will most likely segregate in crosses, giving them great potential in use for discrimination and for the construction of genetic linkage maps.

Genetic similarity using EST-SSR

To test the potential of the EST-SSRs to act as polymorphic markers in genetic studies, allelic data from the 197 EST-SSRs were used to construct a dendrogram (Figure 2) with the studied sugarcane varieties. Genetic similarities based on the Jaccard coefficient were calculated with 1023 polymorphic markers amplified by the EST-SSRs, and they varied from 0.39 to 0.78. The data generated by these markers were not sufficient to differentiate between the two species *Gandacheni* (*S. barberi)* and *Maneria* (*S. sinense)* (0.78), but they could differentiate these two species from *S. officinarum*. The dendrogram shown in this work is in accordance with the origin of the analyzed sugarcane varieties because *S. sinense* and *S. barberi* are usually considered natural hybrids between *S. spontaneum* and *S. officinarum* [44]. In general, the clustering of the analyzed genotypes is in agreement with previous works [18,19,21], validating the new EST-SSRs. The novel EST-SSR markers developed in the present work represent a suitable genetic resource for assessing sugarcane parentage, clonal evaluation, genetic analysis, and to improve direct estimates of functional diversity.

By analyzing the dendogram, observed that most genotypes are grouped because sugarcane species share a narrow genetic base. Commercially cultivated sugarcane is an interspecific hybrid between *S. officinarum* and *S. spontaneum*. About 80% of the genome is contributed by S. *officinarum*, 10% - 15% is contributed by S. *spontaneum* and only 5% - 10% is thought to consist of recombinant

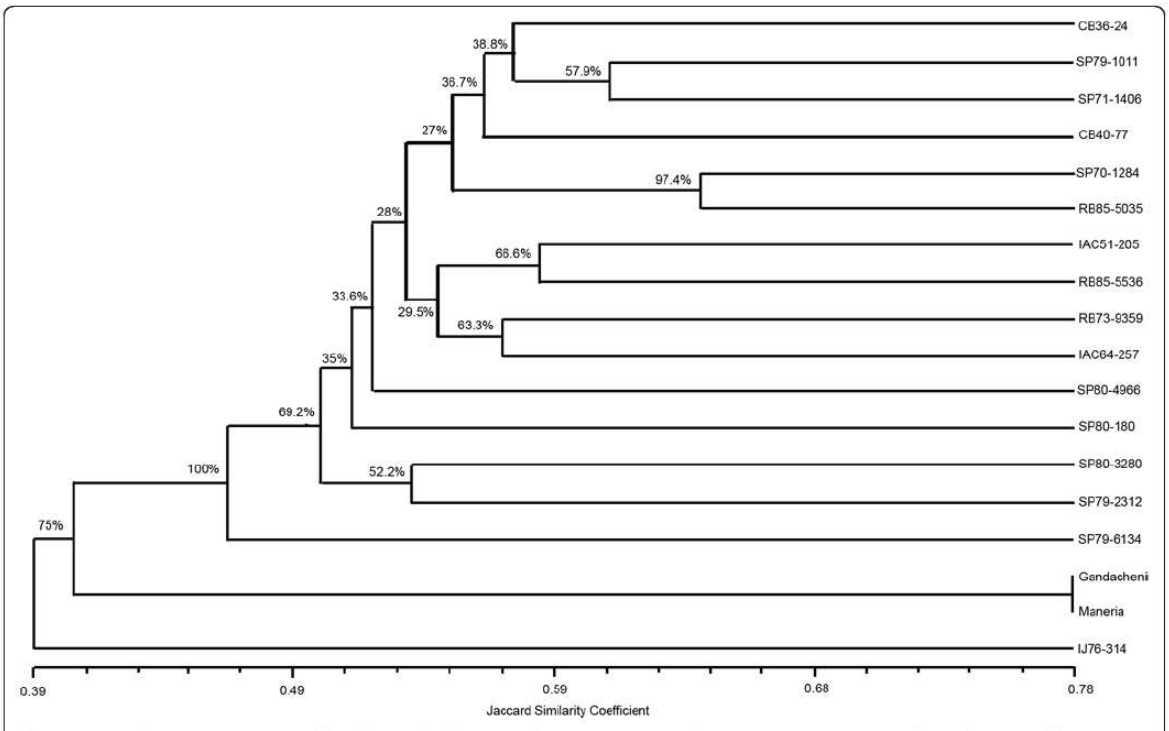


Figure 2 Dendogram representing the relationship between the 18 genotypes of sugarcane, using Jaccard`s similarity coefficients.

chromosomes [45,46]. This may explain why the IJ76-314, that is a *S. officinarum*, cannot be placed in any group and why *Gandacheni* and *Maneria*, both *S. officinarum*, cannot be distinguished themselves, remaining in the same group.

These results corroborate the observed by Lima et al. [22], which observed a tendency for the cultivars to group together with others obtained from the same cross. For example, the *NA56-79* cultivar is associated with crossings that have generated the cultivars *SP79-1011* and *SP71-1406* (Table 1); these cultivars clustered with a similarity coefficient of around 0.60 (Figure 2). Another example includes the genotypes *SP70-1284* and *RB85-5035*. The former is the parent of the cross that generated the latter (Table 1), and they clustered with a similarity coefficient of around 0.65 (Figure 2).

Additional material

Additional file 1: Description of EST-SSR primer pairs characterized in sugarcane. The table presents the markers characterized in sugarcane, as primer sequences, annealing temperatures, number of alleles, expected size, allele range, Polymorphism content and power discrimination.

Additional file 2: Homology and annotation functional EST-SSRs markers. The functional markers were analyzed for their similarity to proteins deposited in GENBANK, further their molecular function and biological process.

Acknowledgements

The authors gratefully acknowledge Renato Vicentini for his constructive and crucial assistance with functional analysis. This work was supported by grants from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (200852197-4) and for giving scholarships to students involved in this work.

Author details

[1]Centro de Biologia Molecular e Engenharia Genética (CBMEG) - Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, CEP 13083-970, Campinas, SP, Brazil. [2]Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana - IAC/Apta, Anel Viário Contorno Sul, Km 321, CP 206, CEP 14.001-970, Ribeirão Preto, SP, Brazil. [3]Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba, SP, Brazil. [4]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6109, CEP 13083-970, Campinas, SP, Brazil. [5]Centro de Tecnologia Canavieira - CTC, Caixa Postal 162, 13400-970, Piracicaba, SP, Brazil.

Authors' contributions

TGM, EAC, MCM and CBCS carried out the molecular genetic studies, participated in the sequence alignment, performed the statistical analysis and drafted the manuscript. HRCAN, KMO and LRP carried out the molecular genetic studies. MM performed the statistical analysis. AAFG and APS conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 28 March 2011 Accepted: 28 July 2011
Published: 28 July 2011

References

1. Carson DL, Botha FC: **Preliminary analysis of expressed sequence tags for sugarcane.** *Crop Sci* 2000, **40**:1769-1779.
2. **Companhia Nacional de Abastecimento.** [http://www.conab.gov.br].
3. Roach BT: **Cytological studies in *Saccharum*. Chromosome transmission in interspecific and intergeneric crosses.** *Proc Int Soc Sugar Cane Technol* 1969, **13**:901-920.
4. Hogarth D: *Genetics of sugarcane. Sugar improvement through breeding* Elsevier Press: Amsterdam; 1987.
5. Silva JAG, Sorrells ME, Burnquist WL, Tanksley SD: **RFLP linkage map and genome analysis of *Saccharum spontaneum*.** *Genome* 1993, **36**:782-791.
6. Burnquist WL: **Os arquitetos da nova cana.** *Revista Pesquisa Fapesp* 2000, **59**:3-4.
7. Jame P, Lagoda PJL: **Microsatellites from molecules to population and back.** *Trends Mol Ecol* 1996, **11**:424-429.
8. Powell W, Machray GC, Provan J: **Polymorphism revealed by simple sequence repeats.** *Trends Plant Sci* 1996, **1**:1021-1026.
9. Wang Z, Weber JL, Zhong G, Tanksley SD: **Survey of plant short tandem DNA repeats.** *Theor Appl Genet* 1994, **88**:1-6.
10. Cato SA, Gardner RC, Kent J, Richardson TE: **A rapid PCR-based method for genetically mapping ESTs.** *Theor Appl Genet* 2001, **102**:296-306.
11. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J: **An optimized protocol for analysis of EST sequences.** *Nucleic Acids Res* 2000, **28**:3657-3665.
12. Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T: **Characteristics, development and mapping of Gossypium hirsutum derived EST-SSRs in allotetraploid cotton.** *Theor Appl Genet* 2006, **112**:430-439.
13. Yu JK, Dake TM, Singh S, Benscher D, Li D, Gill B, Sorrells ME: **Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat.** *Genome* 2004, **47**:805-818.
14. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.).** *Theor Appl Genet* 2003, **106**:411-422.
15. Feng SP, Li WG, Huang HS, Wang JY, Wu YT: **Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (Hevea brasiliensis).** *Mol Breeding* 2009, **23**:85-97.
16. Vettore AL, Silva FR, Kemper EL, Arruda P: **The libraries that made SUCEST.** *Genet Mol Biol* 2001, **24**:1-4.
17. Da Silva JAG: **Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs).** *Genet Mol Biol* 2001, **24**:155-159.
18. Pinto LR, Oliveira KM, Ulian EC, Garcia AA, Souza AP: **Survey in the expressed sequence tag database (SUCEST) for simple sequence repeats.** *Genome* 2004, **47**:795-804.
19. Pinto LR, Oliveira KM, Marconi TG, Garcia AF, Ulian EC, Souza AP: **Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs.** *Plant Breed* 2006, **125**:378-384.
20. Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, Figueira AV, Ulian EC, Garcia AAF, Souza AP: **Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross.** *Molecular Breeding* 2007, **20**:189-208.
21. Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AA, Souza AP: **Characterization of new polymorphic functional markers for sugarcane.** *Genome* 2009, **52**:191-209.
22. Lima ML, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, De Souza CL Jr, Souza AP: **Analysis of genetics similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp.).** *Theor Appl Genet* 2002, **104**:30-38.
23. Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, Silva JAG, Ulian EC, Figueira A, Souza AP: **Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases.** *Theor Appl Genet* 2006, **112**:298-314.
24. Hoisington D, Khairallah M, González-de-Léon D: *Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory* México: CYMMYT; 1994.
25. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ: **Microsatellite markers from sugarcane (*Saccharum* spp) ESTs cross-transferable to *Erianthus* and *Sorghum*.** *Plant Sci* 2001, **160**:1115-1123.

26. Creste S, Tulmann-Neto A, Figueira A: **Detection of single repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining.** *Plant Mol Biol Rep* 2001, **19**:229-306.
27. Mateescu RG, Zhang Z, Tsai K, Phavaphutanon J, Burton Wursten NI, Lust G, Quaas R, Murphy K, Acland GM, Todhunter RJ: **Analysis of Allele Fidelity, Polymorphic Information Content, and Density of Microsatellites in a Genome-Wide Screening for Hip Dysplasia in a Crossbreed Pedigree.** *Journal of Heredity* 2005, **96**:847-853.
28. Tessier C, David J, This P, Boursiquot JM, Charrier A: **Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L.** *Theor Appl Genet* 1999, **98**:171-177.
29. Jaccard P: **Etude comparative de la distribution florale dans une portion des Alpes et du Jura.** *Bulletin Soc Vaud Sc Nat* 1901, **37**:547-579.
30. Sneath PHA, Sokal RR: *Numerical taxonomy* Freeman: San Francisco; 1973.
31. Rohlf FJ: *NTSYSpc: numerical taxonomy and multivariate analysis system, version 2.11X* Applied Biostatistics: New York; 2000.
32. Coelho ASG: *Software BOOD version 3.0* Laboratório Genética Vegetal DBG/ICB: UFG; 2002.
33. Thamarus KA, Groom K, Murrell J, Byrne M, Moran GF: **A genetic linkage map for Eucalyptus globulus with candidate loci for wood, fibre, and floral traits.** *Theor Appl Genet* 2002, **104**:379-387.
34. Schloss S, Mitchell S, White G, Kukatla R, Bowers J, Paterson A, Kresovich S: **Characterization of RFLP probe sequences for gene discovery and SSR development in Sorghum bicolor (L.).** *Moench Theor Appl Genet* 2002, **105**:912-920.
35. Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D'Hont A: **Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane.** *Mol Genet Genomics* 2003, **269**:406-419.
36. Masle J, Gilmore SR, Farquhar GD: **The *ERECTA* gene regulates plant transpiration efficiency in *Arabidopsis*.** *Nature* 2005, **436**:866-870.
37. Zhu YJ, Komor E, Moore PH: **Sucrose accumulation in the sugarcane stem is regulated by the difference between the activities of soluble acid invertase and sucrose phosphate synthase.** *Plant Physiol* 1997, **115**:609-616.
38. Zhang Y, Cheng YT, Bi D, Palma K, Li X: **MOS2, a protein containing G-Patch and KOW Motifs, is essential for innate immunity in Arabdopsis thaliana.** *Curr Biol* 2005, **15**:1936-1942.
39. Cordeiro GM, Pan YB, Henry RJ: **Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm.** *Plant Science* 2003, **165**:181-189.
40. Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P: **Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs).** *Theor Appl Genet* 2004, **109**:800-805.
41. Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D: **A microsatellite map of white clover.** *Theor Appl Genet* 2004, **109**:596-608.
42. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L: **Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species.** *Theor Appl Genet* 2007, **114**:159-372.
43. Cordeiro GM, Taylor GO, Henry RJ: **Characterization of microsatellite markers from sugarcane (*Saccharum* spp.), a highly polyploid species.** *Plant Sci* 2000, **155**:161-168.
44. Price S: **Cytology of Chinese and North Indian sugarcanes.** *Econ Bot* 1968, **22**:155-164.
45. D'Hont A, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N: **Characterization of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics.** *Mol Gen Genet* 1996, **250**:405-413.
46. Aitken KS, Jackson PA, Mcintyre CL: **Construction of a genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage.** *Genome* 2007, **50**:742-756.

doi:10.1186/1756-0500-4-264
Cite this article as: Marconi *et al.*: Functional markers for gene mapping and genetic diversity studies in sugarcane. *BMC Research Notes* 2011 **4**:264.

Additional file 1:

Description of EST-SSR primer pairs characterized in sugarcane. The table presents the markers characterized in sugarcane, as primer sequences, annealing temperatures, number of alleles, expected size, allele range, Polymorphism content and power discrimination.

EST-SSR[a]	Motif	Forward primer (5'-3')	Reverse primer (5'-3')	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB168	(atc)5	CAGCAGCAGCAGTCTTCGTT	GAACTGCGCACCGAAAGA	56.4	2	126	150-170	0.36	0.69
SCB171	(cat)9	GCTTCTTTCTTTTCGTCACACC	TCACCTGACCACACCTCTTTTT	56.4	6	174	240-300	0.79	0.95
SCB174	(agg)5	GGAGATGCTGCGGAGGTGGTT	GCCGCTTCCTCATCATATTCTTCTC	61.0	6	209	200-230	0.78	0.97
SCB180	(cta)5	GGTCCCTGAAGATGAGAGTGAG	CCCATGCATGTAGGTAGGAAT	61.0	2	169	210-220	0.33	0.66
SCB181	(gct)8	GGCGGCTGCTTCTGGGTTTGT	GGAAGCCGAGGAGCACGAGGAT	61.0	9	174	220-300	0.82	1.00
SCB184	(agc)5	GCGTCCACCGGCACCACCTT	CATCCCATCCCGGCACAAGAAGA	62.7	4	153	160-180	0.67	0.86
SCB186	(gat)5	CCTTTGCTTTTTCCCCTTTTC	GGATTACCGTTAGTTCACCCTGTC	58.9	3	214	200-210	0.53	0.75
SCB187	(ttg)5	CGCTCTTCTTTTGTACAGTTCATC	GCTGCTACTCCGACCTTACC	54.2	2	205	250-260	0.28	0.42
SCB188	(gaa)6	CACCGAAGAAACGCCAAAGA	TTCTTCTCACCATCAGCTTCAACAG	58.9	3	201	190-200	0.58	0.81
SCB189	(tca)10	GTAAGGAAGAAGCAACAAACAACAG	GATTCGATGCAACTCTCCTGTAAA	56.4	7	204	280-300	0.79	0.97
SCB190	(tct)8	TTCCTTCTGTCACCATTCATTTG	CCCCTCGATGCTGATTGTTAC	56.4	9	259	250-310	0.84	0.98
SCB191	(gct)5	GCGCCATCAGGGAAGCCAAAAC	GCGCGTGCGAGCAGATGAAC	62.7	3	215	250-260	0.59	0.79
SCB192	(gcg)5	TGGTTGTCCTTCTCCCTTGTGTT	CCCCACTGTCATCCACTCCTTC	58.9	2	165	180-190	0.36	0.74
SCB193	(ttc)6	TTTTGAAGAGAATGTGGACGAG	AGCCAATTACAAACAAAAGGTG	61.0	2	153	160-170	0.37	0.49
SCB194	(atg)7	ATTAACCATCTCGTGTCCCATTCC	TCATCATCTTCGGTGTCCTCTTCA	61.0	3	176	200-230	0.59	0.68
SCB195	(cgc)6	TTCTCGCCGACGCTCACTCC	CTGCCGCCCCTGTTCGTG	61.0	2	338	340-350	0.37	0.39
SCB201	(gag)5	CGCGCCATGATCCTCCTGT	ATTTCCTCCCTTGTCCCCTTCC	62.7	3	170	200-210	0.55	0.53
SCB204	(cat)8	TGCTTCTCGTTGTTATCTTCACC	TGCCAAGTTTACAGAGGAGGAA	61.0	2	150	160-170	0.37	0.37
SCB207	(ctc)7	CGCTTGCCTACCTCGTCTTCTCT	CAGGCGAGCGGGATTGGTA	62.7	8	149	160-200	0.80	0.95
SCB208	(atg)10	TGCAGTTGATTAATTTAGGGTGGTG	CTCAGAAAACAAAATCAAGGTCGTC	61.0	11	194	150-200	0.85	0.97
SCB210	(agg)8	GGAGGCCGAGGACAAGGTAGAG	GCAGCGCACGAGCAGGTC	62.7	2	190	200-250	0.37	0.73
SCB213	(tcc)6	AGCCGTCAGGGGTCAGG	ATTCGATGGAGCCTGAGTGAG	64.7	5	150	160-200	0.74	0.98

EST-SSR[a]	Motif	Forward primer (5′-3′)	Reverse primer (5′-3′)	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB219	(ctt)6	CTCCATGGCCATCGCTGCTCTG	TTCTTTGGTGGTTGGTGGCTGTGG	62.7	6	151	150-180	0.78	0.92
SCB221	(ctc)5	GTGATGGCGCCCTTCTGCT	CCGATCCATGGGGGTTCAC	61.0	7	253	260-310	0.77	0.94
SCB222	(aga)6	TGGGCCGTTGCGTGGGTTGT	GGCGGGCGCAAGAGACTGGA	61.0	4	270	270-290	0.70	0.84
SCB225	(gca)6	GCCTAATGCCATGCCCCAGAG	AGAACCGAACCTGAACTCCGATGTG	56.4	7	208	240-280	0.71	0.93
SCB226	(cgc)5	GCGGAGCAGGCGGAGTG	GCGGTGCCGTGGGGATTA	51.3	5	195	200-230	0.75	0.96
SCB227	(gat)9	GGGGGAGGACGACGACGAGGAC	CACCCTCGCCATCATCCTCCATTT	63.9	6	215	210-240	0.72	0.96
SCB228	(aga)7	CTCCTACGTCTGGCTCCTCCTGTC	GCGTGGTGTCTTGCCTGTGG	64.7	3	213	260-280	0.42	0.54
SCB231	(tca)6	CTCGTCTTCCTCATCGCTGTCTTC	GCGATGAACTGGATGATGACTCTG	61.0	2	161	160-170	0.35	0.69
SCB232	(ggt)5	GGCAAGTGGCAAGGGGAGAT	GATTAATTACCCCAACCGCCAGTC	61.0	2	126	150-160	0.37	0.95
SCB234	(ccg)7	CAACTTCGTACCTACACCAACACC	GAACGAGAACCACGCTGAATAACT	56.4	2	199	210-220	0.28	0.61
SCB236	(tga)5	TCTTTTGGTCCTCATCCTGGTTTGT	TCTGGGCCTCTATATTCTGCTGTGG	63.9	3	177	210-250	0.55	0.80
SCB240	(cag)6	GAGAAGCAGAGCAGCGGGTGGTG	TGATCACCACGCAGCAGAGGAACC	61.0	2	333	320-330	0.25	0.43
SCB241	(aag)6	GGCGCTGTTCCAATTTCCACTG	GCAGCAGCGCAGCCAGAGA	58.9	2	271	340-350	0.36	0.74
SCB243	(tac)6	ATATGGAGCTCCGTCTTCTTGTTA	CCGTAGCTGGGGGTTGAGA	62.7	10	178	180-220	0.87	0.99
SCB246	(gat)6	CACCAGAAACCGATACAACAAAGAC	AGCTCATCAATTCGTCCTATCACAC	61.0	10	142	140-190	0.87	1.00
SCB248	(cat)8	CACACGCATGCCACCAACTC	AAGAGGGGAAAAAGGTTGAGTGTG	61.0	6	330	330-350	0.73	0.82
SCB250	(ccg)5	GCGGGGCGGTGGACGAC	GAGGATGACGACGGTGAGGAAGGAG	62.7	2	147	160-170	0.34	0.53
SCB252	(gct)5	CCCTTTTTGCTGCTTTCTCACTTT	AGTCTCCTCCCGCTGTGATTG	50.4	6	191	190-240	0.75	0.91
SCB253	(tta)5	AGAATGCGATCCTTGACCA	AGCTATACTTGACCTCGCATTAC	54.2	4	282	300-330	0.60	0.82
SCB254	(cat)6	TATGTCGCTCCGTGGGCTCAG	CAAGTTCGGTGGCATGATAGGAGT	51.3	6	213	240-290	0.81	0.93
SCB256	(ggt)5	GGTGTGCGTGGACGGAAGTGAGG	GGCAGCAGGAGGAGGAGCGAGAA	64.7	8	227	230-270	0.84	0.96
SCB257	(gtg)6	CCCTTACCTCGTCGCAATCCTG	AACAAATAACGATCCCGGCACCTAA	64.7	3	233	240-260	0.59	0.06
SCB259	(aac)6	AAACGACAATGGGGGAGAGG	TTATTTGTGCCGTGGTGTTGTAGT	62.7	13	244	240-300	0.86	0.98
SCB262	(tgt)7	TTTTATCTCTAGCCTACCCCAACT	TGCATCAACATAACACATCCAG	62.7	5	258	220-290	0.75	0.84
SCB263	(gtt)6	TTGGCAATTGGAAGGGAAGAT	TGTAGGAGAGGAGGCAACGAC	62.7	10	186	170-200	0.87	0.92
SCB265	(ctc)9	CGCGTTTTCACAGCCCTCTCCA	CCGCCGCCGCCTCTGC	58.9	7	149	170-190	0.83	0.94

EST-SSR[a]	Motif	Forward primer (5′-3′)	Reverse primer (5′-3′)	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB266	(cgc)8	GATGGAGGTGGAGACGATGCTGGAG	CGGAGGAGGCGGGGACGAA	62.7	3	149	130-140	0.50	0.52
SCB267	(ctt)5	CCGGCCACGACGAGGGAGAG	GTGGCGGAACCGGTTGATGAAGT	63.9	5	150	150-160	0.75	0.90
SCB268	(cct)5	CCCCCGGTGGTGAGCAG	CGACGGCGGCGAGGTAG	63.9	5	150	150-180	0.76	0.60
SCB270	(cac)6	CCAACCCCCGCCCCTCCTC	TCGTCGGCGCGGCAGAAGAT	62.7	10	152	130-170	0.85	0.96
SCB271	(gat)6	TTGGTGGAGGGGCTGGATGATGAC	CGCCGGGCCGTACTGACTCTG	62.7	8	153	200-300	0.83	0.96
SCB272	(gtg)5	GTGCAAAGACGAGGATGAGAA	ACTGCCGCGTCAACCAC	62.7	7	153	200-290	0.81	0.63
SCB273	(cag)9	CGACGCCGACATGCCTTCAGT	CTCCACGTTGTCCGCCCATACCT	62.7	8	154	180-290	0.79	0.91
SCB275	(gga)5	GTTCCCAGGATCGTTGTCGTTTTTG	CCTCCTCGTCCACCGCCACTT	61.0	5	155	150-170	0.70	0.63
SCB276	(tgg)6	ACTTGCTGCTAATGATGATGTGG	ACTAGGCGTGTTGGGGTCTG	63.9	4	157	240-300	0.69	0.66
SCB277	(ctt)6	CTCCTTCTTCTTCTTCTTCCTCC	GCTGCCCTAACGCTGCTC	62.7	3	159	250-290	0.55	0.78
SCB279	(tga)6	AGAGGGAGGACAACAACAGG	CTCCAGTCCCAGCATAAAGAT	61.0	7	163	140-170	0.81	0.88
SCB281	(gcc)6	TCCCGTTCGGCCGTTCACCTC	CCGATCATCTTGCCAACCCCTTCA	56.4	8	165	190-300	0.84	0.95
SCB282	(tgg)5	CGCGCTTTGTTTTCTCCTTC	CGCGCACCCCGTCAG	62.7	6	166	170-200	0.81	0.54
SCB283	(ctg)5	GCTTCCGCATGCACTCCTCCAC	GACGCGCCCTGCTCCTACGAC	62.7	5	169	180-230	0.62	0.82
SCB285	(tcc)6	GTAATCATCTTGCCCTCCTCTCCTC	GAACTGCTCACTGGCTCCTCTCA	61.0	9	170	170-250	0.86	1.00
SCB286	(ctc)5	CACGCCGCCGGGAAGGAT	CGAAGGCGGCGGTGGAGAC	62.7	7	170	180-240	0.79	0.82
SCB287	(agg)6	AACGGCGGCGGCATCAGTG	CATCCCGTCGCCGCTCCTCT	61.0	6	125	140-200	0.77	0.90
SCB288	(aag)7	GAGCCGCGCAGCAGCAGAA	GCCATCGTCATCATCATCACAATCC	56.4	6	173	190-270	0.80	0.95
SCB291	(ggc)5	GCCTTCCCGTCGCTCTTACC	CCCCCTCCTCCTCCTCCTCT	62.7	5	177	200-210	0.71	0.86
SCB292	(gcc)6	CCGCGCCGTCCTGGTC	CTAGCACGCCGATTCAAGAGCA	58.9	13	181	160-200	0.88	0.99
SCB293	(ggt)7	GCGGCTTCTGCTCCTGCTCCTC	CTGTCCGCGGCGTCCATCC	62.7	2	184	180-190	0.37	0.11
SCB294	(cgg)11	CCCCGCCGCCGTCCATCT	CGGCCCCGGTCTCCTCTCG	62.7	5	185	190-250	0.72	0.90
SCB295	(ccg)5	GGTGTACTCCATATCCTTGCTGAAA	GCGCGGCGGTGTTGA	50.4	3	188	180-220	0.59	0.71
SCB296	(gca)5	TGCGGCACCTTCTTCCTCAA	GCACGGCGGCTCACCTC	63.9	4	190	200-210	0.67	0.75
SCB297	(tcg)5	CGTGGTTGGCCTTCTCGTC	CCGTCGCCTCTGGTTCG	62.7	7	192	200-290	0.83	0.81
SCB298	(acc)6	GGGAAAGGGGAATGGGGAGAGG	CCGAGCTGGGGAAGAAGACAAGAG	63.9	6	127	180-230	0.77	0.88

EST-SSR[a]	Motif	Forward primer (5′-3′)	Reverse primer (5′-3′)	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB299	(aga)6	CGCCGCCCTTCCGTCTCC	AGCAGCAGCGTCCACATACTCTTCC	50.4	7	194	220-280	0.78	0.92
SCB300	(tct)5	CACCCGGCTTCCTCTCCCAGTCTC	ATCCTCTTCCGCCCTCCCTCTCG	62.7	12	196	190-210	0.88	0.99
SCB301	(tgt)8	TTTGTGTCTCCCTGTTTCTCGTCTC	TTCCCGCAAATGATTCTATGTGG	61.0	5	196	210-260	0.74	0.98
SCB302	(agg)10	ACCCACGCGTCCGAAAACACA	CTCTCACGCCGCTCGCTACACTCT	52.5	2	197	200-220	0.36	0.47
SCB305	(atc)5	TGCCCTGTAAACACGTCCAAAAG	CTCAAACCACCCCCGAAGATAGA	61.0	3	201	230-250	0.57	0.51
SCB306	(cgc)6	CGCCGTCGCTCCTTCT	AAAAACCCTCTACGACCCTAAG	63.9	2	202	210-220	0.25	0.36
SCB307	(cgt)5	GCCCCCTCCTCCCTCGCTTTTC	CACCGCCACCGCCACCTGT	62.7	4	207	200-230	0.59	0.83
SCB309	(agg)8	GGCGGCGAAAACCCTCAAA	CGGCGCCGACCTCCTCTA	62.7	7	208	180-220	0.79	0.97
SCB310	(cag)5	CACCCCCGTTCCCAGCAGAGC	ACAGGAAGGACGCCAGGATGACCA	64.7	4	209	230-260	0.66	0.86
SCB311	(gcg)12	CGACAAGCGGACGAAAAGAGG	CCGCGGTGGAAGGATGGT	62.7	8	211	180-250	0.85	0.99
SCB312	(tcc)7	AGTCCGTCGCCGTAATCATCTTG	GCACCTCCTCCTTTCCTTCCTTATT	64.7	11	215	200-290	0.89	0.99
SCB315	(gct)5	TGCTGGCGCCAAAGACCT	ACTGCTCGCTCTAAACCCCTGTG	61.0	10	218	210-240	0.87	0.94
SCB316	(aga)5	GCTGCGGATGCTGTTAGTGGTGA	TCCTTCCCCTGCTTCTTGCTTTTCT	64.7	11	218	240-300	0.88	0.98
SCB317	(cac)5	CTCCCTTCTCCCCTCTTCCCTTGTC	AGCGCTAGCCCGTCCGTGTTG	62.7	12	219	190-230	0.81	0.94
SCB320	(gag)6	CCCGACGTCGATAAGGAG	CGGGAGGATGTTGCTGAG	58.9	14	225	220-260	0.90	0.99
SCB323	(ctt)5	TGGTCGTCGGAGGTGGAG	CACGGGCGTCAACTGGAT	62.7	7	234	240-300	0.77	0.84
SCB324	(agg)7	GCGAGAGCCAAGAACCAG	GCAGACGGGGCAGAGAA	58.9	6	239	NA	0.79	0.85
SCB329	(ccg)8	CGCCACCGGAAAAACC	CAGACTGCAAGAAAGGAACCA	52.5	14	246	NA	0.91	0.94
SCB330	(agg)6	ATTTCTCAGTCCCTCCTCCTCCTCA	GTAGAAGTCCGTCGCCGTAATCATC	61.0	11	255	200-290	0.88	1.00
SCB332	(tta)5	GCACCAACACCTACAACATCAAGT	CTGGGGAAGGTACGGAACAAG	63.9	5	257	NA	0.70	0.88
SCB336	(cac)6	CGACCCGTCATCCAAAAT	TGCCCTTACCTTCTGTCCA	56.4	8	278	NA	0.78	0.94
SCB337	(tgt)10	CGTGCCTGAACCGTGACC	GGAGCAGCATCAAGAAAAGAAAC	56.4	4	279	NA	0.69	0.77
SCB341	(act)5	GTGTGCGCCTAAAACTGAACCAAC	CAGCAGATGAGCAGACGAAGAAGAA	61.0	3	294	NA	0.45	0.48
SCB343	(cca)7	AGTAGCTCAAGGACAGGGACAG	GAGAGCGGAGAGGATGGTG	63.9	12	300	NA	0.89	0.99
SCB344	(atg)13	ATGCGGCTCGATATCTTGTTGG	CTCCTCCTCCTTCTTATTCACCTCA	61.0	18	126	130-180	0.92	1
SCB345	(cca)6	CCGACGTCGCCGCCAATGTAG	CCACGGGGGCAGCAGGAAGTC	56.4	2	305	NA	0.21	0.30

EST-SSR[a]	Motif	Forward primer (5′-3′)	Reverse primer (5′-3′)	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB347	(cag)5	GCCTTACATCAACAGCATCAACACG	AGAAGATCAAGCTGCAGGAGAACCA	61.0	4	317	NA	0.59	0.65
SCB353	(aag)5	GTCCGGCCCGTCCACACTCC	CGCCGGCGCCAGACTCC	63.9	5	124	NA	0.73	0.45
SCB354	(ctc)5	CGCCGCCCCTGCTGGTG	TGGTGAAAACTACGGCGTGGAACTG	61.0	4	138	NA	0.65	0.79
SCB358	(gag)7	AAACTGCTTGTGGGTGTGG	CGCAGAGGCAGAGGTGAT	62.7	9	168	NA	0.86	0.96
SCB360	(gga)6	CCACAAACGCACCTCAAGAAGTCAA	GGCAGCGGAGGACGACGACA	63.9	10	236	NA	0.85	0.94
SCB361	(gat)7	TGCCACCAGAAACCGATACAAC	AGAGGAGAGGGGAGGGAGAGC	63.9	9	296	NA	0.85	0.92
SCB362	(ctc)7	AACTTCTCTTGCTCCCTCCGTCCTC	GGCGCGTCTGGTCCTGTGG	56.4	3	278	NA	0.58	0.72
SCB365	(cag)5	GTACGGAGCGCAGTGTAGGT	CACCGCCTTGTACTCCAAAT	61.0	10	227	280-190	0.82	0.93
SCB370	(tct)7	CTGGTCGCAGTACTTGATCG	CCTTCCTTGGCCTCTTCTCT	61.0	14	195	310-250	0.89	0.99
SCB371	(gaa)5	GCAACCAGCTAATGAGACGA	GCGTGGAGGAAGAAACAGAG	52.5	12	271	290-250	0.90	0.99
SCB373	(tta)6	TCCATCCATCCTCTCCACTC	CTCTTGGTTCCCCTCACAAA	61.0	13	157	190-150	0.88	1.00
SCB374	(tcg)6	CTCCTTGGCTTGCTCAAAAC	AGCTGCTTCGTATTGCTGGT	61.0	12	278	190-150	0.89	1.00
SCB375	(gca)6	TGTCGGCGGAGTTCTCTACT	AGCTGTAGCCCCAACTGAGA	61.0	8	287	320-250	0.81	0.94
SCB376	(atg)7	AGACGCTATCTGTGGGGAGA	GTATTATGCCCGCTCCTCTG	61.0	7	180	190-160	0.78	0.95
SCB377	(tgg)5	ATAAGCACTGGCCATTCCAC	ACCTTGCTGGTTCAAAGGTG	51.3	2	161	210-200	0.34	0.52
SCB378	(gca)6	ACACAGGTCCCACCTGGATA	CTGAGGAACAGGATGGTGGT	63.9	9	204	230-200	0.88	0.95
SCB379	(gct)6	GTGCACAAGCAATCATCACC	CAAGATGCCAACAGCCAGTA	51.3	2	270	300-290	0.25	0.31
SCB380	(ggt)6	CTCGAGGAGGAACTTGATGG	AAAGCTCACCTCATCGCATC	61.0	6	252	330-290	0.77	0.88
SCB381	(tac)8	TGGAGCTCCGTCTTCTTGTT	GCTAGCCCGTACATTGGGTA	61.0	14	248	260-210	0.88	0.99
SCB384	(aga)7	TTGAAGGTGATGGGGAGAAG	TTACCGGACTCTTCGGTTTG	54.2	6	157	180-160	0.80	0.85
SCB385	(gta)5	ACCGACCAGCTGTTCCTCTT	CCAACTCGGTCATCTCATCA	61.0	11	175	190-170	0.82	0.99
SCB388	(gga)6	AGTACGCTAACCGTCCGTTG	AGTGGGAGTCCTCCTCTTCG	51.3	6	228	330-290	0.74	0.94
SCB393	(ccg)7	TCTATGCAGATCCACCACGA	AAGGGAATTACCCCAAGCAG	62.7	12	197	210-190	0.86	0.84
SCB395	(ctc)5	ACGACCTCGTCGAATCCTTT	TCCTCCTGAATGCACTCCTC	61.0	5	207	220-200	0.71	0.88
SCB397	(gtg)8	CTGTGGGACTACTCGCCTTC	GGAGGAGGCTGTGATTAGCA	61.0	6	235	230-210	0.80	0.95
SCB400	(gtg)5	GCCTGGTGAATCCAACTGAT	CTCCACGGCCATTAGGATAG	61.0	10	183	210-160	0.84	0.95

EST-SSR[a]	Motif	Forward primer (5′-3′)	Reverse primer (5′-3′)	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB401	(cta)5	AAGAAACCTTGCCTGGGTCT	GTTCATCATCTCCAGCAGCA	61.0	4	285	290-280	0.66	0.67
SCB402	(cat)5	AACCTGCAATAGCCCATGAC	TGACATCTGAGTTGCCTTCG	61.0	2	219	320-310	0.33	0.52
SCB403	(cat)5	GTTCTCCTCATCAGCTGCAA	AGAGCCTCAGCTCTGCTCTG	50.4	4	252	330-320	0.57	0.82
SCB404	(gct)7	TGTGAGGATCGGAGTACGTG	GGCATATACACACGCCACTG	61.0	9	197	210-180	0.84	0.92
SCB405	(atg)7	TGGCCGTAGGGAAAGATATG	TCGCAGCATCTTCATCAGAC	50.4	2	268	280-270	0.38	0.00
SCB407	(tca)6	ACGCAACTCGCTGAAGGTAG	GTCCTTGGTGGAATCTGACG	61.0	2	270	190-150	0.36	0.42
SCB408	(gag)5	GACTTGGACGACGACACCAG	CGCATAGCTAGCCAAGATCC	61.0	10	258	310-230	0.84	0.89
SCB409	(gcc)9	ACCGGAAAAGCCAGAGACTT	GTCGTCCATGTGATGTGCTC	51.3	8	182	190-120	0.79	0.98
SCB413	(aca)6	CCATGTCTCTTTGGTTCCTG	TCTAGCCTACCCCAACTTGC	61.0	7	156	170-150	0.78	0.89
SCB416	(tgt)8	TAGCCTACCCCAACTTGCTT	GGCCCTCCATATAAGTTCTG	51.3	2	163	170-160	0.26	0.29
SCB417	(att)5	GGATTGCCGAGGATCTACAG	ATACAGGCGACCATGGATTC	61.0	6	193	210-180	0.66	0.78
SCB418	(tgt)5	GTTTGCCTACCCCAACTTGC	TAATTCCTGCACCTCCTTGC	61.0	2	169	180-170	0.35	0.54
SCB419	(tgt)8	CCCTTTGTGTGTCTCCCTGT	CCAGAGGTGGTAAACCTTGC	61.0	4	159	170-150	0.69	0.84
SCB422	(gat)7	GCCTGCTATTTGCTCGATTC	TTTGGAGCACCCTCTCAATC	61.0	9	220	240-210	0.84	0.97
SCB423	(ttg)5	CCATGTGGCTTCCTGAAACT	ACAGGCACTTCAAGGGAAGA	61.0	15	233	300-250	0.88	0.99
SCB427	(acc)8	GGACATGCTGCTCCCTACAT	AGGAGGACTGGTGGTTGAGG	62.7	8	211	220-190	0.80	0.63
SCB428	(gaa)9	CTGGCCTCAAGAGGAAACTG	TTCTGCGTACTGGTCGATGA	61.0	9	302	300-260	0.86	0.98
SCB429	(gag)5	ATGCCCATGAGAAGAAGCTC	GAAGAAGCGCTCCAAACAAG	61.0	9	189	200-180	0.80	0.91
SCB430	(cac)6	TTCCCTCCAAACTTCCTCCT	GTCCGACATCTCATCCTCGT	61.0	9	294	220-180	0.82	0.86
SCB431	(gaa)5	CAAGGACAAGGTGCTCGTTT	CGGGTACGGATATGGGTATG	61.0	12	253	270-220	0.86	0.98
SCB432	(cca)5	GCGTCCGATAGCATGTAGGT	ATGTGTCTCAGGGTGTGCAA	54.2	4	256	330-260	0.63	0.29
SCB433	(cct)5	CCGGACTCCCACTCTCCAAG	CATGGTGCCGCCTAGTGTAG	61.0	6	156	250-210	0.74	0.73
SCB434	(tgt)6	TCTAGCCTACCCCAACTTGC	TCTCTTGAAACGGCACCTTC	64.7	12	259	310-250	0.89	0.99
SCB435	(atg)5	GCTTCTGCTGCTACGAGACC	GCACTTGCACGACCAGACTA	61.0	6	201	220-190	0.78	0.92
SCB436	(gag)5	AGTACGCCTGAGTCCTGACG	AGGTGCAAGGGCTGATAGAA	61.0	13	181	280-180	0.90	0.96
SCB437	(cgg)12	ATAGAACTTCCCCTCTGACC	GACTGTTACGGGGTCTCTTC	50.0	5	291	310-290	0.69	0.90

EST-SSR[a]	Motif	Forward primer (5'-3')	Reverse primer (5'-3')	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB438	(ggt)5	TCCGATCCTACTGGTGGGTA	TTGTAGTTGGCTGCACCGTA	61.0	5	299	320-290	0.74	0.81
SCB439	(cca)5	CCCATACCCAAACTGCCATA	TTGCTGTTCAGTTCGCTACG	61.0	9	277	290-260	0.86	0.87
SCB440	(cac)9	GGAAGAGGAGGCTTCGAGAT	CAGATCCCCGATACAACCAG	61.0	7	286	170-130	0.76	0.88
SCB441	(acc)7	GATGCTTCTCCTCCGTCTTG	AGACAGAGCAAGGCGATCAT	61.0	8	165	170-120	0.80	0.95
SCB442	(agg)5	ACACACACCTCGTCTTGCAG	CCATGTTTGGGCTATCACCT	61.0	6	285	310-280	0.69	0.92
SCB443	(aga)5	GTGAAGCCGGTGAAGGAG	GTCTCAGCTCCCTTGCTGTC	63.9	3	233	240-220	0.59	0.29
SCB445	(cag)6	GAACCAAGAACAGGAAGC	GATCGTCGTCTAACCTTGTG	61.0	2	192	190-180	0.37	0.37
SCB446	(cag)6	GCTACCAAGCAGGCTATCCA	CAGCATCCTTTCCAGAAACC	61.0	4	187	190-180	0.58	0.82
SCB447	(gcc)10	CCCCCTGCCTCCTACAGT	GTGGTGGGTGTCAGTTTGC	61.0	6	281	330-310	0.78	0.99
SCB449	(cta)5	AAGAAACCTTGGCTGGGTCT	CCCTCAACGAAAGGTAGGTG	50.0	2	251	260-250	0.37	0.11
SCB450	(gta)5	CCACCTACTACCTCGCTTGC	TGTCCGTGTTCCACTAGACG	61.0	9	182	190-170	0.83	0.94
SCB451	(tag)5	AGACTGTAGCGCTTGAACCA	GCAGTCTGCTTGATGTGTGAG	61.0	9	169	200-160	0.76	0.92
SCB452	(agt)5	CTGGCAACCTACCAGCAAAT	TTTCGCTCTTGGTCGACTCT	61.0	4	288	300-290	0.59	0.63
SCB453	(tat)6	GCCTAGCGAATCCAAATC	GAAATAGCCGCAACTTCG	50.0	4	197	220-190	0.47	0.67
SCB456	(atc)5	GCAGCTAGCAGCAACAAGAA	TGGCTACAACTTCGGCTGAT	61.0	7	173	190-170	0.64	0.69
SCB457	(atc)10	TCCACCTCTTCCCAAAACAC	CGGTCAGAGTGGGTTCACTT	61.0	8	154	170-140	0.81	0.82
SCB458	(atc)5	CCGTACGTGCATCTCTCTCA	TCATGCAGTGCAGAGGAATC	61.0	5	279	280-250	0.65	0.67
SCB459	(gga)5	TTTAGCGACGGCGAGTCTAT	CCTCGCTCTCCTCTCCTCTC	61.0	2	150	180-140	0.23	0.29
SCB462	(gat)5	TGGGAATGAGGAGGAAGATG	TCCTGCATGTAAGCAACCAC	61.0	7	250	270-230	0.80	0.85
SCB464	(tcc)6	GTTCCTCAGCCCTAGCTTCC	CGGATCGCCACCAGTAAGTA	61.0	4	189	190-180	0.47	0.58
SCB466	(gac)6	AGCAGCAGAATGAAGCGTGCCA	CGACAGGGAGAGGGAGAGCAGG	63.9	16	204	240-180	0,92	1
SCB467	(ttc)5	GTTTCCAACCCGTCTGCCTGCC	ATTGGGCGCAAAGCAACCCTGT	58.9	5	275	300-260	0,69	0,75
SCB469	(cct)5	GTCGTCATCGCCACCACCAGTT	CGACATGTACCTCGCGCCATCC	63.9	4	237	330-310	0,73	0,75
SCB472	(gag)5	ACAAACAGGCCTCGGGTCCTCA	CGCTTGTGGATGTCGCCGATGA	65.0	3	183	200-180	0,6	0,6
SCB474	(cca)5	GATCAAGGGCGCCAAGGGTTCC	CGCCTTTGTGCTCCTCTGCAACT	62.7	9	152	170-140	0,85	0,63
SCB476	(ggc)7	CTTCTCGGTTTCTGACTTGTTCCG	TGGCTGGGGTTGACGTAGATGG	61.0	7	265	290-260	0,68	0,71

EST-SSR[a]	Motif	Forward primer (5'-3')	Reverse primer (5'-3')	Tm (°C)	A[b]	E. size (bp)[c]	A. range(bp)[d]	PIC[e]	DP[f]
SCB477	(ggc)6	GACCCCGTCATACCTGCACCGA	CGGTCGATCGGAAGCTTTCCCC	58.9	9	192	230-190	0,85	0,99
SCB478	(aga)5	CACGGCATGGCAATCAGGTGTG	GGGTTCTTATCCCTCGTATCCAGTC	58.9	4	199	210-180	0,64	0,78
SCB480	(acc)7	AGGCCAGCTCAAGGCCAACAGA	CTTTACCGCGAACAGCTCCCCG	52.5	4	242	260-220	0,78	0,75
SCB483	(cca)8	TCCGCTGCTCGTCATGGAGTCA	ACGGAATCGAAGAAGCACCGCC	61.0	6	263	280-240	0,83	0,66
SCB484	(act)5	ACCACGGGACATGCGATGCG	AGCTAGAAGCAGAAGTGGAGTGCT	61.0	8	196	230-180	0,84	0,76
SCB485	(tta)5	GGACCTAATGTGAACCACAG	GAAGCTTTCTGGCTGTCTGA	50.4	7	194	220-180	0,77	0,86
SCB486	(tgt)6	GCCTACCCCAACTTGCTTGGGAC	GGGGCAAAAATGGCCTTCACGTCT	52.5	9	225	250-160	0,8	0,96
SCB487	(cac)6	AGCCTCCATCTCCGACACACCT	AGCCTGCCTAGTAGTTCGCCGT	52.5	5	279	290-250	0,61	0,77
SCB489	(aga)17	CCTGACCAGGAGACTGAAGAGCCA	GACAGCAGGCATCGCAGGTTGA	58.9	8	243	260-210	0,8	0,91
SCB491	(tgt)7	ACACACGAGACATGACAACTGCC	CAACATCCTGCATTGGTGAAGTGC	62.7	4	252	260-240	0,6	0,75
SCB492	(acg)8	GTGGGCCTAATGGGGTGCCAAG	TTGCCAGCTGCGCTTTTCCTCA	62.7	8	286	310-270	0,85	0,96
SCB502	(ctt)5	GGAGGAAGCAGGCAGGAGAGGG	AGGGGCAGGATGAGCGAGGAAG	63.9	6	203	240-180	0,82	0,69
SCB504	(agt)5	AGGCACGCTTGCTAGCGGAGAT	GCGGTGGTCGGGAGGAACAAAT	62.7	3	208	220-200	0,42	0,55
SCB518	(cac)5	GATGCTGGAACCACTGGAGCCG	CAAGTACAGACCGCGAACCCCG	65.0	3	208	230-210	0,51	0,66
SCB524	(ccg)5	CCTCCTTTCCTCTCCCTCCCGC	TGCCACAGATAGGACCCTCGGC	50	6	268	NA	0,77	0,79
SCB525	(gta)9	GTCCATGTCCATCCATGATCCA	GACTGACCAGCAACCGTCAAAA	62.7	9	348	330-280	0,83	0,93
SCB526	(tta)7	GCTGAGAGAGACATTCACATGC	GCTGCTGGTGGGGAAAATAA	58.9	4	179	180-170	0,61	0,75
SCB527	(taa)9	TGTCAAGTGGAGGCAGCTGCTATGT	GGAGGACAGGGATGGCTACCTGC	65.0	8	261	270-230	0,81	0,94
SCB528	(atc)8	AGCGCAACAGAGAGGGAGAGGG	AGCATGGGAGAGAAAGAGGAGAGGT	62.7	11	248	280-220	0,88	0,98

[a] SCB is a trinucleotide primer pair;
[b]A: allele number;
[c] This is the expected size as calculated by the Primer Select software (LaserGene versions 5.01/5.02);
[d] Allele range estimated by linear regression;
[e]Polymorphism information content;
[f]Discriminant power.

Additional file 2:

Homology and annotation functional EST-SSRs markers. The functional markers were analyzed for their similarity to proteins deposited in GENBANK, further their molecular function and biological process.

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 186	L-allo-threonine aldolase [EU961815]	0	BP-3	MF-7
SCB 187	brassinosteroid insensitive 1-associated receptor kinase 1 [NM_001154503]	0	BP-2; BP-3	MF-5; MF-1; MF-13
SCB 188	SAR DNA-binding protein [NM_111378]	7E-119		
SCB 189	putative ubiquitin-specific protease UBP27 [AY133770]	9E-27	BP-3	MF-4; MF-9
SCB 190	WRKY1 [NM_001112367]	0	BP-2	MF-1
SCB 191	aspartyl protease family protein [NM_001161138]	3E-10	BP-13	
SCB 192	DNA-3-methyladenine glycosylase I [EU975955]	8E-135	BP-3	MF-4
SCB 193	putative amino acid permease gene [GQ252800]	9E-15	BP-6	MF-15
SCB 194	Minghui 63 receptor kinase (MRKd) [DQ355952]	2E-155		
SCB 195	hypothetical protein [XM_002466252]	0		
SCB 201	carboxyl-terminal proteinase [EU963164]	4E-176		
SCB 204	hypothetical protein [XM_002462405]	1E-104		
SCB 207	Receptor protein kinase CLAVATA1 precursor [XM_002531953]	5E-13		MF-1
SCB 208	DNA-binding with one finger1 (dof1) [NM_001112239]	2E-26	BP-2	MF-1
SCB 209	serine/threonine-protein kinase SAPK4 [EU974118]	0		
SCB 210	ring finger protein [XM_002531237]	3E-39	BP-3	MF-6; MF-1

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 213	phosphoglycerate mutase-like protein [EU956232]	2E-76	BP-3	MF-2
SCB 218	soluble acid invertase [AF062735]	0	BP-16	MF-4
SCB 219	arsenite transport subunit B [NM_001154652]	0		MF-15
SCB 221	SAUR family protein (SAUR24) [XM_002320183]	2E-10	BP-1	MF-1
SCB 222	hypothetical protein [XM_002447224]	3E-50		
SCB 225	ethylene-insensitive3-like 1 protein [NM_001158563]	0		MF-13
SCB 226	protein binding protein [NM_001156822]	1E-90		
SCB 227	nucleotide binding protein [NM_001153645]	1E-58		MF-1
SCB 228	Retrosat 1 retrotransposon and Ty3-Gypsy type Retrosat 2 [AF111709]	9E-07		
SCB 231	SDA1 family protein [NM_101186]	3E-133	BP-5	MF-1
SCB 232	homeodomain leucine zipper protein [AF145726]	1E-79	BP-2	MF-1; MF-13
SCB 234	ATSWI3B (SWITCH SUBUNIT 3) [NM_128921]	5E-43		
SCB 236	nodulin protein [NM_001158210]	1E-112	BP-6	MF-15
SCB 240	hypothetical protein [NM_001175846]	6E-23		
SCB 241	rac-like GTP-binding protein 5 [EU968843]	1E-148	BP-6; BP-9	MF-1
SCB 243	Group peroxidase 16 precursor protein [EF990901]	2E-158	BP-1; BP-7	MF-8
SCB 246	hypothetical protein [XM_002450508]	0		
SCB 248	ubiquitin thiolesterase [NM_147880]	5E-83	BP-3	MF-4
SCB 250	hypothetical protein [XM_002444735]	4E-83		
SCB 252	ATP binding protein [EU956876]	5E-143		MF-1

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 253	ATPase family protein [NM_100170]	0		MF-1
SCB 254	VHS and GAT domain-containing protein [NM_129397]	4E-84	BP-6	
SCB 256	bZIP transcription factor (bZIP111) [DQ787052]	2E-12	BP-2	MF-1
SCB 257	no hit	-		
SCB 259	Sorghum bicolor clone BAC 75D9 [AY661659]	3E-17		
SCB 262	no hit	-		
SCB 263	hypothetical protein [XM_002453018]	4E-19		
SCB 265	plastidic phosphate translocator-like protein1 [EU965602]	2E-69		
SCB 266	hypothetical protein [XM_002441011]	6E-47		
SCB 267	hypothetical protein [XM_002437132]	2E-48		
SCB 268	protein binding protein [NM_001154397]	9E-73		
SCB 270	stress responsive protein [EU956106]	3E-66	BP-2	MF-1; MF-13
SCB 271	hypothetical protein [XM_002440411]	6E-65		
SCB 272	plant-specific domain TIGR01568 family protein [NM_001153443]	3E-44		
SCB 273	ccr4-associated factor putative [XM_002523816]	5E-18	BP-2	MF-13
SCB 275	armadillo-repeat containing protein [EU957148]	3E-45		
SCB 276	hypothetical protein [XM_002488881]	1E-93		
SCB 277	MYB transcription factor (MYBAS1) [EU670236]	6E-60		MF-1
SCB 279	hypothetical protein [XM_002458804]	1E-65		
SCB 281	transport protein particle, component Bet3 family protein [NM_125188]	1E-111	BP-6	

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 282	hypothetical protein [NM_001150840]	7E-27		
SCB 283	putative proline-rich protein [AY096387]	1E-11	BP-2	MF-13
SCB 285	putative auxin efflux carrier (PIN1b) [DQ836240]	5E-77		
SCB 286	putative tyrosine phosphatase [AF272978]	3E-117	BP-11; BP-3	MF-4
SCB 287	uridylate kinase [EU965854]	1E-156	BP-3	MF-1
SCB 288	ubiquitin-conjugating enzyme 23 (UBC25) [XM_002882888]	1E-32	BP-3	MF-6
SCB 291	ubiquitin-specific protease 25 (UBP25) [XM_002884996]	7E-13	BP-3	MF-4
SCB 292	hypothetical protein [NM_001152536]	3E-24		
SCB 293	dehydration responsive element binding protein [NM_001156946]	3E-34	BP-2	MF-1
SCB 294	hypothetical protein [XM_002460181]	4E-143		
SCB 295	putative kinase [AF360190]	8E-171		
SCB 296	putative protein kinase [AF466200]	5E-38		
SCB 297	hypothetical protein [XM_002454582]	1E-61		
SCB 298	homeodomain leucine zipper protein (Oshox4) [AF145728]	1E-52	BP-2	MF-13; MF-1
SCB 299	hypothetical protein [XM_002459270]	9E-27		
SCB 300	rac-like GTP-binding protein 6 [EU960626]	0	BP-9	MF-1
SCB 301	leucine-rich repeat protein (LRR2) [EF555120]	1E-21		MF-1
SCB 302	putative auxin efflux carrier (PIN1b) [DQ836240]	9E-51		
SCB 305	leucine-rich repeat transmembrane protein kinase 1 (ltk1) [AF023164]	0	BP-3	MF-1; MF-5
SCB 306	RNA (guanine-9-) methyltransferase domain containing, putative [XM_002533353]	2E-88	BP-3	MF-14

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 307	hypothetical protein [XM_002451430]	3E-50		
SCB 309	nucleic acid binding protein [XM_002525476]	9E-11	BP-2	MF-1
SCB 310	chloroplast phytoene synthase (Y1) [AY455286]	5E-41	BP-3	MF-14
SCB 311	polyadenylate-binding protein [XM_002510992]	9E-38		MF-1
SCB 312	putative auxin efflux carrier (PIN1b) [DQ836240]	0		
SCB 315	ASL5 mRNA for ASYMMETRIC LEAVES2-like 5 protein [AB473838]	2E-48		
SCB 316	hypothetical protein [NM_001175157]	1E-168		
SCB 317	PAN1 (pan1) [FJ231525]	8E-44	BP-6	
SCB 320	hypothetical protein [XM_002447675]	6E-66		
SCB 323	fasciclin-like arabinogalactan protein 7 [NM_001154093]	0	BP-12	
SCB 324	putative nucleic-acid binding protein (MOS2) [AJ311811]	3E-20	BP-4	MF-1
SCB 329	Pi21 gene for Pi21 protein [AB430854]	5E-18		
SCB 330	putative auxin efflux carrier (PIN1b) [DQ836240]	3E-78		
SCB 332	mapk-alt-1 mitogen-activated protein kinase (MAPK) [EU280318]	2E-31	BP-3	MF-5; MF-1
SCB 336	putative nucleic-acid binding protein (MOS2) [DQ202264]	4E-45	BP-4	
SCB 337	putative MADS-domain transcription factor [AJ850302]	0	BP-2	MF-13
SCB 341	RNA-binding protein [NM_001153713]	7E-33		MF-1
SCB 343	transferase mRNA [EU958141]	2E-60		MF-14
SCB 344	hypothetical protein [XM_002440411]	4E-60		
SCB 345	hypotetical protein [EU970828]	1E-86		

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 347	potyvirus VPg interacting protein [NM_001154424]	0		
SCB 353	hypotetical protein [EU973478]	2E-23		
SCB 354	transferase, transferring glycosyl groups [XM_002532362]	2E-34		MF-14
SCB 358	no hit	-		
SCB 360	hypothetical protein [XM_002460057]	2E-44		
SCB 361	hypothetical protein [AY596525]	1E-75		
SCB 362	hypothetical protein [NM_001147848]	0.004		
SCB 365	no hit	-		
SCB 370	VAMP [NP_001150642]	3E-96	BP-17	MF-11
SCB 371	no hit			
SCB 373	ring zinc finger [NP_001151265]	3E-23	BP-3	MF-1; MF-6
SCB 374	nitrate transporter [BAC83856]	2E-33	BP-6	MF-15
SCB 375	methyl binding [NP_001105174]	2E-53		MF-1
SCB 376	hypotetical protein [XP_002447452]	7E-54		
SCB 377	RNA recognition motif containing protein [CBH19571]	3E-67	BP-3; BP-14	MF-1
SCB 378	ATP binding protein [XP_002526287]	3E-56		MF-4
SCB 379	putative RNA binding protein [AAG59664]	1E-109	BP-10; BP-6	MF-1
SCB 380	phosphoglycerate kinase [ABI74567]	0.0	BP-3	MF-5
SCB 381	peroxidase 16 precursor [NP_001152697]	7E-154	BP-1; BP-7	MF-1; MF-8
SCB 384	SAR DNA binding protein [BAA31260]	0.0		

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 385	cinnamoyl-CoA reductase [CAA13176]	0.0	BP-3	MF-2; MF-1
SCB 388	putative Clp protease ATP-binding subunit ClpX1 [BAD15818]	7E-27	BP-6; BP-3	MF-4; MF-14; MF-1
SCB 393	zinc finger (C2H2 type) family protein [NP_566182]	2E-30	BP-7	MF-3; MF-1
SCB 395	tyrosyl-tRNA synthetase [NP_001148699]	0.0	BP-3	MF-1; MF-6
SCB 397	zinc finger, C2H2 type family protein [NP_001151975]	6E-62	BP-7	MF-3; MF-1
SCB 400	hypothetical protein [XP_002466649]	3E-135		
SCB 401	pnFL-2 [ACG35001]	2E-65		MF-12
SCB 402	lipid binding protein, putative [XP_002523536]	0.0		
SCB 403	syntaxin 32 [NP_001149760]	7E-149	BP-6	MF-15
SCB 404	FPF1 [NP_001151726]	6E-43	BP-9; BP-1	
SCB 405	kelch repeat-containing protein [NP_568723]	7E-40		MF-1
SCB 407	FACT complex subunit SPT16 [NP_001105557]	7E-12	BP-13	MF-9
SCB 408	receptor-like protein kinase-like protein [AAO72615]	2E-14	BP-3	MF-5; MF-1
SCB 409	hypothetical protein [XP_002447767]	5E-05		
SCB 413	bile acid sodium symporter [ACG42083]	4E-09	BP-6	MF-15; MF-10
SCB 416	casein kinase II alpha subunit [BAB21589]	2E-101	BP-3	MF-5; MF-1
SCB 417	multifunctional protein [AAT38103]	0.0		
SCB 418	PHD-finger family protein [ABA91581]	2E-41	BP-2	MF-1; MF-13
SCB 419	HVA22-like protein c [ACG36039]	6E-37		
SCB 422	hypothetical protein [XP_002443307]	5E-48		

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 423	vacuolar protein sorting 26 [ACG47931]	1E-11	BP-6	
SCB 427	hypothetical protein [XP_002442368]	2E-39		
SCB 428	hypothetical protein [XP_002463328]	2E-35		
SCB 429	CLE family OsCLE306 protein [BAF91618]	1E-08		
SCB 430	homeobox-leucine zipper protein HAT7 [NP_001150309]	1E-34	BP-2	MF-1; MF-13
SCB 431	extensin [NP_001152552]	4E-16	BP-8	MF-11
SCB 432	hypothetical protein [NP_001140516]	3E-28		
SCB 433	putative potyviral helper component protease-interacting protein 2 [BAD45803]	2E-21	BP-6	MF-15
SCB 434	serine/threonine-protein kinase AFC3 [ACG35326]	4E-41	BP-3	MF-5
SCB 435	hypothetical protein [ACG29480]	2E-11		
SCB 436	hypothetical protein [ACG24888]	7E-14		
SCB 437	serine/threonine-protein kinase N1 [NP_058871]	4.6	BP-3	MF-5
SCB 438	hypothetical protein [XP_002463153]	4E-06		
SCB 439	hypothetical protein [NP_001169794]	7E-11		
SCB 440	hypothetical protein [XP_002448081]	1E-08		
SCB 441	no hit	-		
SCB 442	no hit	-		
SCB 443	heavy metal-associated domain containing protein [ABF95901]	3E-18	BP-6	MF-1
SCB 445	no hit	-		
SCB 446	no hit	-		

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 447	hypothetical protein [EEC71447]	0.058		
SCB 449	pnFL-2 [ACG35001]	4E-06		
SCB 450	no hit	-		
SCB 451	no hit	-		
SCB 452	no hit	-		
SCB 453	no hit	-		
SCB 456	VQ motif family protein [NP_001147304]	6E-09		
SCB 457	no hit	-		
SCB 458	no hit	-		
SCB 459	hypothetical protein [XP_002445708]	1E-34		
SCB 462	nucleolin [NP_001149300]	3E-24	BP-3	MF-1; MF-6
SCB 464	no hit	-		
SCB 466	Zinc finger, C2H2 type [AAX95339]	2,00E-13	BP-3	MF-1
SCB 467	hypothetical protein [NP_001168922]	9E-140		
SCB 469	NADH-ubiquinone oxidoreductase 39 kDa subunit [NP_001149069]	0	BP-3	MF-5
SCB 472	brushy protein [XP_002517217]	3,00E-68	BP-10	MF-1
SCB 474	pherophorin like protein [NP_001152156]	1,00E-51		MF-11
SCB 476	pantoate--beta-alanine ligase [ACG42174]	4,00E-144	BP-3	MF-6
SCB 477	bZIP family transcription factor [NP_564269]	1,00E-56	BP-3	MF-13
SCB 478	oxidoreductase, zinc-binding dehydrogenase family protein [ABF93481]	8,00E-87		MF-1

EST-SSR*	Best-matched protein[a] [GenBank Accession[b]]	BLAST E-value	Biological Process	Molecular Function
SCB 168	hypotetical protein [AAG13074]	4E-11		
SCB 171	plant-specific domain TIGR01627 family protein [NP_001147380]	5E-43		
SCB 174	dnaK-type molecular chaperone hsp70 - rice [CAA47948]	3E-11	BP-2	MF-1; MF-13
SCB 180	lyrata ubiquitin family protein [XM_002871569]	2E-41		
SCB 181	ESK1 (ESKIMO 1) [NM_115457]	3E-138	BP-2	
SCB 184	cytokinin response regulator2 (crr2) [NM_001111379]	1E-63		
SCB 480	NPKL1 [ACH99693]	3,00E-43		
SCB 483	retrotransposon protein [ABB47941]	2,00E-23		
SCB 484	hypothetical protein [XP_002453128]	2,00E-14		
SCB 485	hypothetical protein [XP_002452863]	9,00E-43		
SCB 486	uridylate kinase plant [XP_002526510]	9,00E-19		MF-1
SCB 487	sugar transport protein 5 [ACG44242]	0	BP-6	MF-12
SCB 489	putative translation initiation factor IF-2 [AAT44302]	3,00E-22	BP-9	MF-13
SCB 491	methylmalonate semi-aldehyde dehydrogenase [AAC03055]	2,00E-04	BP-3	MF-8
SCB 492	hypothetical protein [XP_002460882]	1,00E-81		
SCB 502	Associated with HOX family protein [ABA91569]	3,00E-143		
SCB 504	hypothetical protein [XP_002449910]	8,00E-28		
SCB 518	no hits	-		
SCB 524	hypothetical protein [XP_002448610]	8,00E-09		
SCB 525	no hits	-		
SCB 526	no hits	-		
SCB 527	no hits	-		
SCB 528	no hits	-		

[a] EST-homology: annotation of the best homology identified by BLASTX and available in the SUCEST database.
[b] GenBank identification.
* These EST sequences were from the SUCEST database and are not in GenBank; therefore. they do not have a GI number.

ARTIGO V

"SNPs como uma nova ferramenta no mapeamento genético em espécies poliploides e aneuploides: exemplos obtidos com cana-de-açúcar"

Marconi, Thiago G.; Rodrigues, Renato; Mollinari, Marcelo; Margarido, Gabriel R.A.; Vicentini, Renato; Bundock, Peter; Henry, Robert J.; Pinto, Luciana R., Garcia, Antonio A. F., de Souza, Anete P.

Artigo não submetido

SNPs como uma nova ferramenta no mapeamento genético em espécies poliploides e aneuploides: exemplos obtidos com cana-de-açúcar

Marconi, Thiago G.[1]; Rodrigues, Renato[3]; Mollinari, Marcelo[3]; Margarido, Gabriel R.A.[3]; Vicentini, Renato[1]; Bundock, Peter[2]; Henry, Robert J.[2]; Pinto, Luciana R.[5], Garcia, Antonio A. F.[3], de Souza, Anete P.[1,4]

[1]Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, Campinas, SP CEP 13083-875, Brazil; email: anete@unicamp.br, thimarconi@gmail.com; [2]Centre for Plant Conservation Genetics, Southern Cross University, Lismore NSW 2480, Australia.; email: robert.henry@scu.edu.au; [3]Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, Piracicaba, SP CEP 13400-970, Brazil; email: aafgarci@esalq.usp.br; [4]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, CP 6010, Campinas, SP CEP 13083-875, Brazil; email: anete@unicamp.br; [5]Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana, IAC/Apta, Anel Viário Contorno Sul, Km 321, CP 206, Ribeirão Preto, SP, CEP 14001-970, Brazil; email: lurossini@iac.sp.gov.br

Resumo

As espécies cultivadas de cana-de-açúcar possuem um grande número cromossômico, geralmente variando de 100 a 130, além de serem aneuplóides. O marcador molecular mais utilizado para construção de mapas genéticos nestas espécies é o SSR que possui a desvantagem de funcionar com um marcador dominante neste caso. Por este fato, somente marcadores SSR segregando em dose única (1:1 e 3:1) ou em dose dupla (7:2) podem ser

inseridos ao mapa, ainda assim de forma bastante limitada. Em estudos preliminares realizados com cana-de-açúcar usando SNPs como marcador codominante evidenciou-se que o maior número de polimorfismos encontrado não é em dose única, mostrando a necessidade do desenvolvimento de marcadores moleculares e técnicas de análise para que todos os tipos de segregação possam ser incorporados ao mapa genético, aumentando a sua cobertura e resolução e consequentemente aumentando o poder das buscas por características quantitativas. O presente trabalho apresenta uma forma de obter marcadores SNPs de tal forma que todos os tipos de doses sejam obtidos, o que pode melhorar os mapas genéticos. A incorporação de marcadores com segregação 1:2:1 é apresentada como exemplo das vantagens que tal abordagem propicia.

Introdução

As variedades cultivadas de cana-de-açúcar são híbridos interespecíficos entre *S. officinarum* (2n=8x=80 cromossomos) e *S. spontaneum* (2n=40-128, x=8 cromossomos)(Ming et al., 2001). Após as hibridações iniciais, feitas pelos primeiros melhoristas da espécie, foi realizada uma série de retrocruzamentos com *S. officinarum*, no processo chamado de nobilização, que reduziu a contribuição de *S. Spontaneum e* aumentou o teor de açúcar (Oliveira et al., 2007), mantendo as característica de rusticidade. Dessa forma, a maioria dos cromossomos da cana cultivada são derivados de *S. officinarum*, sendo que cerca de 15% parece ter derivado de *S. spontaneum*, de acordo com estudos empregando hibridação genômica *in situ* (GISH) (D'Hont et al., 1996). Nota-se que o genoma do chamado complexo Saccharum é altamente poliplóide e aneuploide. (Jannoo et al., 2004). Assume-se portanto que os estudos realizados neste complexo genoma possam ser de grande utilidade não só para a cana-de-açúcar, mas também para os poliploides de forma geral.

Entre as ferramentas disponíveis para a realização de estudos genéticos, merecem destaque os marcadores moleculares, que têm se mostrado particularmente úteis para auxiliar programas de melhoramento em diversas culturas (Francia et al., 2005). Esses marcadores vêm sendo utilizados em cana-de açúcar para estudos de diversidade genética (Lima et al., 2002), obtenção de mapas genéticos (Aitken et al., 2005, 2007; Raboin et al., 2006; Garcia et al., 2006; Oliveira et al., 2007) e mapeamento de QTLs (Quantitative Trait Loci) (MCintyre et al., 2006; AL-Janabi et al., 2006; Piperidis et al., 2008) e até mesmo mapeamento associativo, embora num primeiro momento com maior ênfase em estudos de desequilíbrio de ligação (Jannoo et al., 1999; Raboin et al., 2008). Trabalhos utilizando marcadores moleculares, especialmente para a construção de mapas genéticos de ligação e mapeamento associativo, são importantes para aumentar o entendimento da estrutura do genoma da cana-de-açúcar, bem como da arquitetura genética dos caracteres quantitativos. Tais estudos podem diminuir o tempo de obtenção de variedades melhoradas, através da identificação de marcadores associados a caracteres de interesse econômico ou locos de características quantitativas (QTLs, ou, mais recentemente, QTNs, quando SNPs estão disponíveis).

Para obtenção de mapas genéticos em cana de açúcar, usualmente são utilizados marcadores SSRs, principalmente devido a sua robustez, relativa abundância no genoma, natureza multialélica e herança codominante, muito embora em cana-de-açúcar tal marcador tenha herança dominante(Oliveira et al., 2009). São incluídos nos mapas apenas alelos segregando em dose única presente em um dos genitores (segregando na progênie na proporção de 1:1), dose única presente nos dois genitores (proporção de 3:1) e até mesmo em dose dupla em alguns casos (proporção de 7:2) (MCintyre et al., 2009; Alwala, 2010). O uso de marcadores moleculares em dose única para o mapeamento genético em poliploides foi um grande avanço nas últimas décadas (Wu et al., 1992), porém devido a estreita base de marcadores utilizados, já que somente determinados tipos de segregação podem ser usados,

não é possível a obtenção de um mapa com uma boa cobertura do genoma. Para tanto, é necessário um avanço em relação ao desenvolvimento de outros marcadores, em que seja possível a distinção de outras doses em poliploides, para que assim outras segregações possam ser utilizadas no mapeamento genético.

Os marcadores SNPs (*Single Nucleotide Polymorphisms*) possuem um grande potencial para suprir a demanda de marcadores moleculares integrados a mapas genéticos em cana-de-açúcar. Estes constituem variações na sequência de DNA que ocorrem quando um único nucleotídeo na sequência do genoma é alterado. Esses polimorfismos, juntamente com as deleções e inserções, são responsáveis pela maior parte da variação na maioria dos organismos,(Cho et al., 1999; Rafalski, 2008) além de serem amplamente distribuídos pelo genoma, sendo mais abundantes em regiões não transcritas e em regiões que flanqueiam microssatélites (Mogg et al., 2002; Bundock e Henry, 2004). Por serem abundantes em muitas espécies de plantas como milho (Tenaillon et al., 2001), aveia (Kanazin et al., 2002) e arroz (Yu et al., 2002), os SNPs têm sido muito utilizados como marcadores, criando perspectivas para o diagnóstico de doenças, a seleção assistida por marcadores (MAS), a construção de mapas genéticos de alta resolução e a caracterização varietal (Batley et al., 2003) sendo considerado o sistema de marcadores mais atraente desenvolvido até o momento (Gupta et al., 2001). Além de apresentarem uma distribuição ampla, esses marcadores apresentam facilidades para high-throughput (Giancola et al., 2006; Masouleh et al., 2009), sendo que as tecnologias e abordagens desenvolvidas para estudos em humanos (Tanaka, 2005; Frazer et al., 2007; Burton et al., 2007) têm facilitado a aplicação desses marcadores nos mais diversos organismos (Ferguson et al., 2007; Kim, 2007; Bain et al., 2007).

Diferentes metodologias vêm sendo utilizadas para identificação de SNPs em diferentes espécies. Um exemplo é a utilização de um banco de dados de sequências expressas (Buetow et al., 1999; Suliman-Pollatschek et al., 2002; Batley et al., 2003; Kota et al., 2003). Graças,

principalmente, ao programa SUCEST, há uma grande disponibidade de ESTs em cana-de-açúcar, que podem ser utilizados como base para busca *in silico* de SNPs de interesse. Poucos trabalhos foram realizados para avaliar o potencial do uso de banco de ESTs para o busca de SNPs em cana-de-açúcar, porém todos os resultados obtidos se mostram bastante promissores (Grivet et al., 2003; MCintyre et al., 2006; Cordeiro et al., 2006). Outra metodologia utilizando sequenciamento de nova geração também vem sendo utilizada com sucesso na identificação de SNPs em cana-de-açúcar (Bundock et al., 2009).

Diferentes técnicas de genotipagem de SNPs também vêm sendo desenvolvidas, como a utilização de pirosequenciador (Cordeiro et al., 2006) e de espectrômetro de massa (Sequenom MassArray)(Bundock et al., 2009). Devido ao alto nível de ploidia da cana-de-açúcar, o pirosequenciador não apresenta resolução suficiente para detecção de alelos com frequências extremamente baixas, como nas razões 1:9 ou 1:10, sendo mais indicado o uso de espectrometria de massa (Cordeiro et al., 2006; Bundock et al., 2009), o qual fornece um resultado quantitativo podendo ser inferido o número de cópias do alelo presente entre os homólogos de um determinado indivíduo, possibilitando a analise do marcador SNP como um marcador codominante, algo até hoje raro para organismos poliploides.

Diante desse cenário este trabalho teve por objetivo o desenvolvimento de marcadores moleculares codominantes (SNPs) em cana-de-açúcar a partir de sequências expressas (SUCEST) que foram genotipados utilizando espectrometria de massa (Sequenom MassArray), possibilitando a incorporação de marcadores com segregação 1:2:1 a um mapa genético previamente construído com marcadores SSRs, exemplificando as vantagens dessa nova abordagem. Com isso, será possível aumentar sensivelmente a cobertura do genoma e a busca por QTLs.

Material e Métodos

População de mapeamento

A população utilizada é derivada do cruzamento biparental de variedade pré-comerciais, IACSP95-3018 (genitor feminino) e IACSP93-3046 (genitor masculino). Os genitores são contrastantes para as características de produção de colmos e resistência a ferrugem. A progênie derivada deste cruzamento possui 220 indivíduos e é mantida no Centro de Cana de Ribeirão Preto do Instituto Agronômico de Campinas (IAC)

Extração DNA

O DNA genômico foi extraído a partir do meristema da cana-de-açúcar, utilizando a metodologia descrita por AL-Janabi et al. (1999). A quantificação do DNA foi feita utilizando o corante fluorescente PicoGreen dsDNA quantitation kit (Invitrogen) e a leitura da fluorescência foi realizada utilizando o leitor de placas MWGt Sirius HT-TRF microplate reader (MWG). Após a quantificação, todas as amostras foram diluídas em água para a concentração final de 5ng/ul.

Identificação de marcadores SNPs

Seleção dos genes candidatos e sequências

Dada a abundância de seqüências ESTs disponíveis no banco de dados do SUCEST, neste trabalho foi utilizado esta fonte de informação para a identificação e desenvolvimento de SNPs como possíveis marcadores genéticos. Os genes para os quais estão sendo desenvolvidos os marcadores moleculares do tipo SNP foram selecionados a partir do projeto SUCAST (Sugarcane Signal Transduction), que tem como objetivo identificar ESTs para componentes de transdução do sinal no genoma da cana-de-açúcar (Rocha et al., 2007). A determinação dos genes de

interesse utilizou as mais diversas fontes de informações disponíveis, como por exemplo, perfis de expressão gênica determinados por experimentos de microarrays, análise de domínios funcionais e similaridade com ontologias moleculares e celulares (http://sucest-fun.org).

Identificação dos SNPs

Aliada a esta busca, foram desenvolvidas ferramentas de bioinformática, capazes de avaliar in silico a ocorrência de SNPs nos *clusters* selecionados. Isto foi feito em colaboração com o Prof. Dr. Renato Vicentini do Laboratório de Bioinformatica e Biologia de Sistema localizado no Centro de Biologia Molecular e Engenharia Genética – UNICAMP. A predição de SNPs considera a ocorrência de mais de um nucleotídeo em uma única posição do *cluster*, sendo considerada apenas as bases com qualidade *phred* >= 20. A metodologia é baseada em modificações do software *QualitySNP* (Tang et al., 2006), buscando assim determinar corretamente a variação alélica. Para tal, foram realizadas buscas por potenciais SNPs, reconstrução de haplótipos, identificação de parálogos e identificação de SNPs não sinônimos.

Genotipagem dos SNPs

Para a genotipagem dos SNPs foi utilizado o sistema Sequenom® MassARRAY® (espectrômetro de massa MALDI-TOF) que envolve uma reação de PCR utilizando um par de *primers* específicos a região de interesse contendo o SNP, seguido por outra reação de PCR utilizando um *primer* especifico adjacente ao loco SNP e a enzima Iplex para a amplificação de um nucleotídeo utilizando dideoxinucleotídeos com massas modificadas, utilizando a espectrometria de massa é possível diferenciar a massa dos diferentes alelos do SNP com uma alta precisão (Gabriel et al., 2009). Todas as etapas foram realizadas na Southern Cross

University, localizada em Lismore, Austrália.

Classificação dos SNPs

Todos os locos SNPs genotipados foram classificados em relação a segregação utilizando testes de segregação usuais, baseados na estatística qui-quadrado e considerando a realização de múltiplos testes (Garcia et al., 2006). Vale lembrar que procedimentos mais sofisticados serão adotados quando todas as segregações forem classificadas (etapas posteriores do presente trabalho).

Resultados e discussão

Seleção e Identificação dos SNPs a partir do SUCEST

Foram selecionadas 1500 sequências únicas (*clusters*) do SUCEST para posterior alinhamento e procura de SNP, devido à ocorrência de diferencial de expressão evidenciado (Rocha et al., 2007; Papini-Terzi et al., 2009). Todos os *reads* utilizados na montagem dos *clusters* selecionados foram alinhados e com a ajuda do programa *QualitySNP* (Tang et al., 2006) foi possível a identificação de mais de 3000 potenciais SNPs. A abundância relativa de SNPs não foi muito grande quando comparada com outro trabalho (Bundock et al., 2009); possivelmente, isto aconteceu devido a pequena quantidade de sequências utilizadas no alinhamento para a procura dos locos SNPs. O bando de dados do SUCEST foi construído a partir do sequenciamento de uma variedade de cana-de-açúcar (SP80-3280) utilizando a tecnologia *sanger*, que produz sequências relativamente longas porém com poucas repetições, ao contrário de outros trabalhos que utilizam sequenciamento de nova geração que produzem

sequências pequenas porem altamente repetidas, possibilitando uma melhor precisão no alinhamento e detecção dos locos SNPs (Bundock et al., 2009).

Desenvolvimento e genotipagem dos locos SNPs

Foi possível o desenvolvimento de *primers* para a amplificação de 942 locos SNPs agrupados em conjuntos de 10 locos por reação, totalizando 95 *multiplexes* que foram submetidos à validação. A partir dos 942 locos SNPs desenvolvidos foi possível a amplificação de 790 (84%), os quais foram testados no restante da progênie. Deste total foram obtidos 245 locos SNPs avaliados na população de mapeamento com um bom padrão de amplificação.

Avaliação das segregações encontradas

Com o conjunto de dados obtido a partir da genotipagem de 245 locos SNPs, foram feitas análises das segregações de cada loco. Foi constatado que a maioria das segregações não são em dose única (1:1 e 3:1), mas sim segregações em altas doses. Este resultado é surpreendente devido ao grande número de trabalhos que sugerem que a maior parte dos marcadores moleculares em cana-de-açúcar possuem segregação em dose única (Alwala, 2010). Isto evidencia que apenas uma pequena porção do genoma está sendo amostrado pelos marcadores moleculares quando estes são restritos apenas a marcadores em dose única. Esta parte se tornou a mais laboriosa do trabalho pelo fato da indisponibilidade de ferramentas para interpretação dos dados gerados, sendo que o programa disponibilizado pelo pacote Sequenom (*Typer Analyser*) é indicado apenas para análise de organismos diplóides, que obviamente não é o caso da cana-de-açúcar. Sendo assim novas ferramentas de análise estão em processo de desenvolvimento pelo grupo para que a interpretação dos resultados obtidos com o

equipamento seja adequada quando forem genotipados organismos poliploides.

Em análises preliminares realizadas é possível a observação de segregações nunca antes utilizadas em mapeamento genético em cana-de-açúcar, que é o caso da segregação em dose única com a distinção das três classes (1:2:1). Este tipo de segregação não pode ser observada utilizando marcadores SSRs, pois apesar do marcador SSR ser codominante na maioria das espécies diploides, em cana-de-açúcar ele é utilizado como um marcador dominante devido a sua natureza altamente poliploide e aneuploide impossibilitando a quantificação do número de cópias com essa técnica. No caso de SNPs, pode-se observar como exemplo na Figura 1, claramente podemos ver a distinção de 3 classes, uma classe onde o parental 2 (P2) possui apenas um dos alelos juntamente com 25% da progênie, outra classe mais ao centro representando os indivíduos heterozigotos (50%) e a terceira classe acompanhada pelo parental 1 (P1) e mais alguns indivíduos da progênie (25%) que se diferenciam da classe central por possuir algumas cópias a mais de um dos alelos.

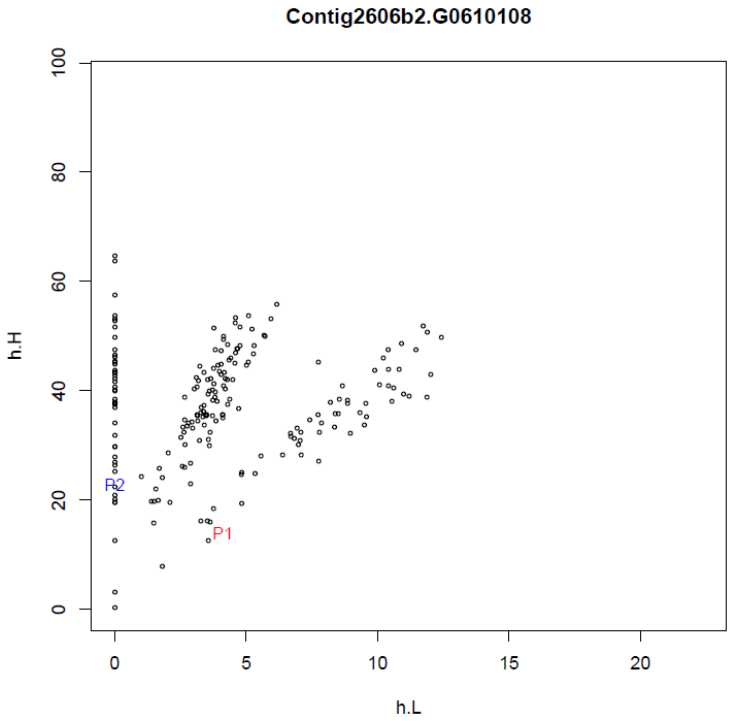


Figura 1. Representação gráfica utilizando todos os indivíduos da progênie e os genitores em destaque organizados em um plano cartesiano com os eixos representando a massa de cada alelo, sendo que, h.H é o alelo de maior massa e h.L é o alelo de menor massa.

Mapeamento genético utilizando SNPs

A partir dos locos desenvolvidos foram avaliados no total 42 SNPs, foram encontrados 31 SNPs segregando em dose única, dentre os quais 25 segregando na proporção 1:1 e 6 segregando na proporção 1:2:1. O restante dos locos apresenta um padrão de segregação multi-dose e estão servindo como base para o desenvolvimento de novos modelos estatísticos para análise desse novo tipo de dado em organismos poliploides e aneuploides. Utilizando um mapa genético prévio como base, construído apenas com marcadores SSRs e AFLP, os 42 locos SNPs segregando em dose única foram inseridos ao mapa utilizando o programa *OneMap* (Margarido et al., 2007). Dentre estes 42, 31 foram incluídos no mapa por apresentarem

evidências de ligação com marcadores do mapa.

Como resultado, podem-se observar grandes mudanças no arranjo dos grupos de ligação e mesmo dentro dos grupos de homologia. Os locos conseguiram um melhor posicionamento dentro do grupo de ligação e uma melhor distância entre os locos pode ser calculada. Algumas mudanças são exemplificadas na Figura 2 em que os grupos de ligação 47 e 3 (mapa prévio) juntaram-se formando um grande grupo de ligação 3A com a adição de apenas 3 SNPs, dois em dosagem 1:1 e um em dosagem 1:2:1, além do rearranjo de alguns blocos de marcadores. Outro destaque foi a retirada do grupo 3 de uma das pontas, que antes era considerada parte do grupo, formando agora o grupo 3B.

Outro exemplo de rearranjo e junção de grupos de ligação com a adição de locos SNPs também é observado no grupo 42 e 33 (mapa prévio) onde um novo grupo agora chamado 38 é formado, totalizando 290,9 cM. Essa mudança deu-se com a adição de apenas 3 locos SNPs, dois segregando na proporção 1:1 que permitiram a junção de 2 outros grupos prévios e um SNP segregando na proporção 1:2:1 que permitiu o rearranjo de alguns marcadores previamente inseridos ao mapa.

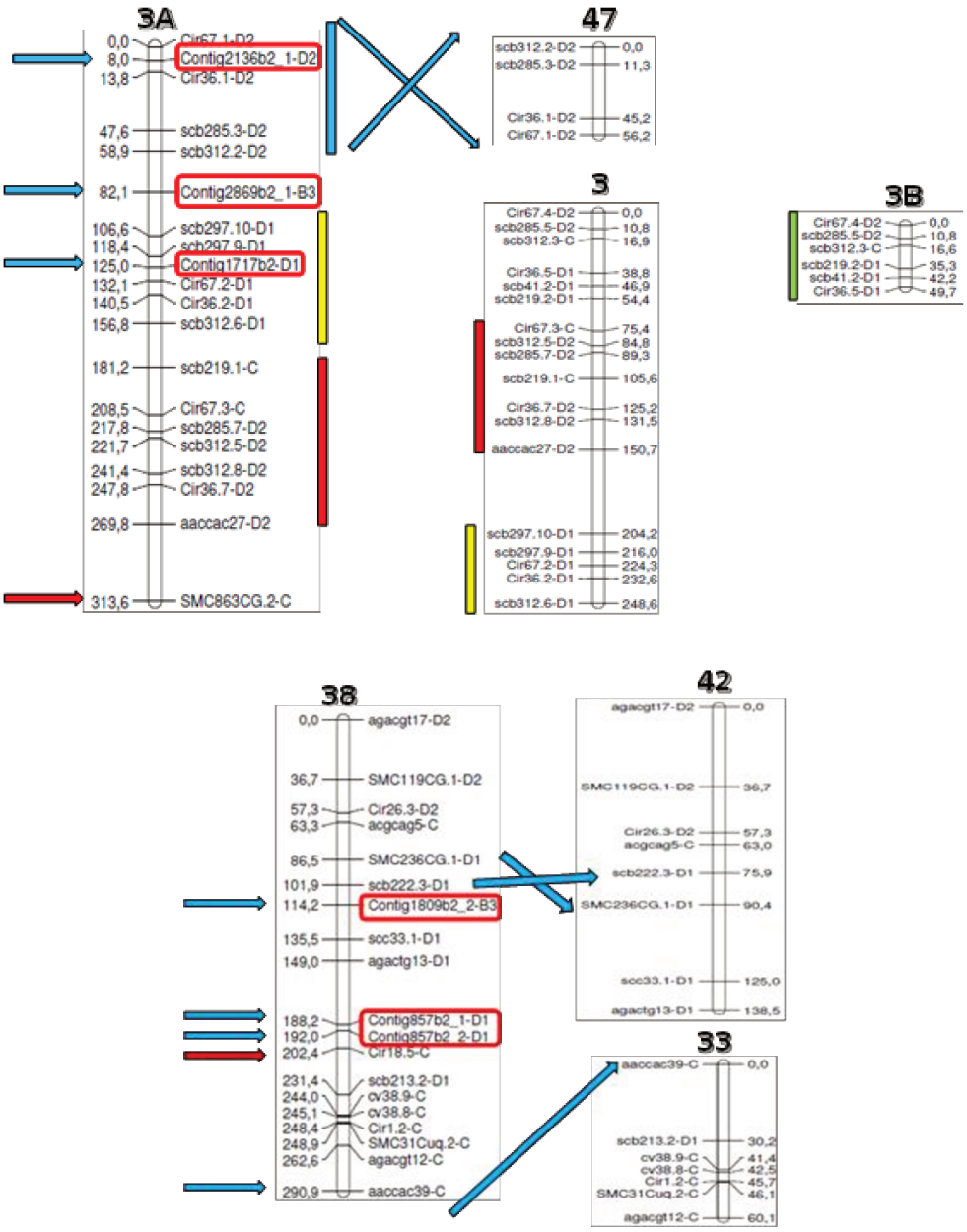


Figura 2. Representação de grupos de ligação formados por marcadores SSRs, AFLPs e SNPs (em destaque), as setas representam as principais mudanças que ocorreram assim como a junção de grupos de ligação. Os SNPs estão indicados com círculos vermelhos.

Conclusões

A partir do desenvolvimento de oligonucleotídeos para identificar locos SNPs em cana-de-açúcar e genotipagem utilizando espectrometria de massa foi possível a obtenção de informações nunca antes disponíveis em organismos poliploides e aneuploides, especificamente para cana-de-açúcar, no qual existe a possibilidade da quantificação do número de cópias de cada alelo presente no loco, ou seja, um marcador molecular realmente codominante para espécies poliploides. Com isso aumentamos a gama de marcadores que podem ser utilizados em mapeamento genético nessas espécies. Por serem mais informativos, os SNPS permitirão uma melhor resolução e saturação nos mapas genéticos que permitirão a localização com mais precisão de QTLs, já que os únicos marcadores que podem ser utilizados atualmente, marcadores em dose única, não conseguem distinguir as classes de heterozigotos, diferente dos SNPs que torna possível a distinção das dosagens alélicas e a distinção das diferentes classes de segregação encontradas em poliploides. Os resultados aqui apresentados mostram claramente os novos horizontes que esta tecnologia permite.

Bibliografia

AITKEN, K. S.; JACKSON, P. A.; MCINTYRE, C. L. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. **TAG. Theoretical and applied genetics.**, v. 110, n. 5, p. 789-801, 2005.

AITKEN, K. S.; JACKSON, P. A.; MCINTYRE, C. L. Construction of a genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. **Genome Research**, v. 50, n. 8, p. 742-756, 2007.

AL-JANABI, S. M.; PARMESSUR, Y.; KROSS, H.; DHAYAN, S.; SAUMTALLY, S.; RAMDOYAL, K.; AUTREY, L. J. C.; DOOKUN-SAUMTALLY, A. Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. **Molecular Breeding**, v. 19, n. 1, p. 1-14, 2006.

ALJANABI, S.; FORGET, L.; DOOKUN, A. An improved and rapid protocol for the isolation of polysaccharide-and polyphenol-free sugarcane DNA. **Plant Molecular Biology Reporter**, v. 17, n. 3, p. 281–282, 1999.

ALWALA, S.; K., C. A. K. Molecular Genetic Linkage Mapping in Saccharum: Strategies, Resources and Achievements. In: Robert J. Henry; C. Kole (Eds.); **Genetics, genomics and breeding of sugarcane**. 1st ed. CRC Press, p.69-96, 2010.

BAIN, J. M.; TAVANTI, A.; DAVIDSON, A. D.; JACOBSEN, M. D.; SHAW, D.; GOW, N. A. R.; ODDS, F. C. Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. **Journal of clinical microbiology**, v. 45, n. 5, p. 1469-1477, 2007.

BATLEY, J.; BARKER, G.; O´SULLIVAN, H.; EDWARDS, K. J.; EDWARDS, D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. **Plant physiology**, v. 132, n. 1, p. 84-91, 2003.

BUETOW, K. H.; EDMONSON, M. N.; CASSIDY, A. B. Reliable identification of large numbers of candidate SNPs from public EST data. **Nature Genetics**, v. 21, n. 3, p. 323-325, 1999.

BUNDOCK, P. C.; ELIOTT, F. G.; ABLETT, G.; BENSON, A. D.; CASU, R. E.; AITKEN, K. S.; HENRY, R. J. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. **Plant biotechnology journal**, v. 7, n. 4, p. 347-354, 2009.

BUNDOCK, P. C.; HENRY, R. J. Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. **TAG. Theoretical and applied genetics.**, v. 109, n. 3, p. 543-551, 2004.

BURTON, P. R.; CLAYTON, D. G.; CARDON, L. R.; et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. **Nature genetics**, v. 39, n. 11, p. 1329-1337, 2007.

CHO, R. J.; MINDRINOS, M.; RICHARDS, D. R.; SAPOLSKY, R. J.; ANDERSON, M.; DRENKARD, E.; DEWDNEY, J.; REUBER, T. L.; STAMMERS, M.; FEDERSPIEL, N.; THEOLOGIS, A.; YANG, W. H.; HUBBELL, E.; AU, M., CHUNG, E. Y.; LASHKARI, D.; LEMIEUX, B.; DEAN, C.; LIPSHUTZ, R. J.; AUSUBEL, F. M.; DAVIS, R. W.; OEFNER, P. J. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. **Nature genetics**, v. 23, n. 2, p. 203-207, 1999.

CORDEIRO, G. M.; ELIOTT, F.; MCINTYRE, C. L.; CASU, R. E.; HENRY, R. J. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. **TAG. Theoretical and applied genetics,** v. 113, n. 2, p. 331-343, 2006.

D'HONT, A.; GRIVET, L.; FELDMANN, P.; RAO, S.; BERDING, N.; GLASZMANN, J. C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Molecular & general genetics : MGG**, v. 250, n. 4, p. 405-413, 1996.

FERGUSON, B.; STREET, S. L.; WRIGHT, H.; PEARSON, C.; JIA, Y.; THOMPSON, S. L.; ALLIBONE, P.; DUBAY, C. J.; SPINDEL, E.; NORGREN, R. B. Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). **BMC genomics**, v. 8, p. 43, 2007.

FRANCIA, E.; TACCONI, G.; CROSATTI, C.; BARABASCHI, D.; BULGARELLI, D.; DALL'AGLIO, E.; VALÈ, G. Marker assisted selection in crop plants. **Plant Cell Tissue and Organ Culture**, v. 82, n. 3, p. 317-342, 2005.

FRAZER, K. A; BALLINGER, D. G.; COX, D. R.; et al. A second generation human haplotype map of over 3.1 million SNPs. **Nature**, v. 449, n. 7164, p. 851-861, 2007.

GABRIEL, S. ZIAUGRA, L.; TABBAA, D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. **Current protocols in human genetics**, v. Chapter 2, Unit 2.12, 2009.

GARCIA, A. A. F.; KIDO, E. A; MEZA, A. N.; SOUZA, H. M. B.; PINTO, L. R.; PASTINA, M. M.; LEITE, C. S.; DA SILVA, J.; FIGUEIRA, A.; ULIAN, E. C.; SOUZA, A. P. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **TAG. Theoretical and applied genetics**, v. 112, n. 2, p. 298-314, 2006.

GIANCOLA, S.; MCKHANN, H. I.; BÉRARD, A.; CAMILLERI, C.; DURAND, S.; LIBEAU, P.; ROUX, F.; REBOUD, X.; GUT, I. G.; BRUNEL, D. Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. **TAG. Theoretical and applied genetics.**, v. 112, n. 6, p. 1115-1124, 2006.

GRIVET, L.; GLASZMANN, J-C.; VINCENTZ, M.; DA SILVA, F.; ARRUDA, P. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. **TAG. Theoretical and applied genetics**, v. 106, n. 2, p. 190-197, 2003.

GUPTA, P. K.; ROY, J. K.; PRASAD, M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, v. 80, n. 4, p. 524–535, 2001.

JANNOO, N.; GRIVET, L.; DAVID, J.; D'HONT, A; GLASZMANN, J-C. Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. **Heredity**, v. 93, n. 5, p. 460-7, 2004.

JANNOO, N.; GRIVET, L.; DOOKUN, A.; D'HONT, A.; GLASZMANN, J. Linkage disequilibrium among modern sugarcane cultivars. **TAG Theoretical and Applied Genetics**, v. 99, n. 6, p. 1053–1060, 1999.

KANAZIN, V.; TALBERT, H.; SEE, D.; DECAMP, P.; NEVO, E.; BLAKE, T. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). **Plant molecular biology**, v. 48, n. 5-6, p. 529-537, 2002.

KIM, S.; MISRA, A. SNP genotyping: technologies and biomedical applications. **Annual review of biomedical engineering**, v. 9, p. 289-320, 2007.

KOTA, R.; RUDD, S.; FACIUS, A.; KOLESOV, G.; THIEL, T.; ZHANG, H.; STEIN, N.; MAYER, K.; GRANER, A. Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). **Molecular genetics and genomics : MGG**, v. 270, n. 1, p. 24-33, 2003.

LIMA, M. L. A.; GARCIA, A. A. F.; OLIVEIRA, K. M.; MATSUOKA, S.; ARIZONO, H.; DE SOUZA, C. L.; DE SOUZA, A. P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). **TAG Theoretical and Applied Genetics**, v. 104, n. 1, p. 30-38, 2002.

MARGARIDO, G. R. A.; SOUZA, A. P.; GARCIA, A. A. F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, v. 144, n. 3, p. 78-79, 2007.

MASOULEH, A. K.; WATERS, D. L. E.; REINKE, R. F.; HENRY, R. J. A high-throughput assay for rapid and simultaneous analysis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI-TOF mass spectrometry. **Plant biotechnology journal**, v. 7, n. 4, p. 355-363, 2009.

MCINTYRE, C. L. G., M. MONKS, T.; BONNETT, GRAHAM D. The Complex Genetic Structure of Sugarcane Limits Identification of Additional SNP-Defined Simplex Alleles in Microsatellite Loci. **Tropical Plant Biology**, v. 2, n. 3-4, p. 133-142, 2009.

MCINTYRE, C.L.; JACKSON, M.; CORDEIRO, G. M.; AMOUYAL, O.; HERMANN, S.; AITKEN, K. S.; ELIOTT, F.; HENRY, R. J.; CASU, R. E.; BONNETT, G. D. The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. **Molecular Breeding**, v. 18, n. 1, p. 39-50, 2006.

MOGG, R.; BATLEY, J.; HANLEY, S.; EDWARDS, D.; O'SULLIVAN, H.; EDWARDS, J. Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. **TAG. Theoretical and applied genetics.**, v. 105, n. 4, p. 532-543, 2002.

OLIVEIRA, K. M.; PINTO, L. R.; MARCONI, T. G.; MARGARIDO, G. R. A.; PASTINA, M. M.; TEIXEIRA, L. H. M.; FIGUEIRA, A. V.; ULIAN, E. C.; GARCIA, A. A. F.; SOUZA, A. P. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. **Molecular Breeding**, v. 20, n. 3, p. 189-208, 2007.

OLIVEIRA, K. M.; PINTO, L. R.; MARCONI, T. G.; MOLLINARI, M.; ULIAN, E. C.; CHABREGAS, S. M.; FALCO, M. C.; BURNQUIST, W.; GARCIA, A. A. F.; SOUZA, A. P. Characterization of new polymorphic functional markers for sugarcane. **Genome Research**, v. 52, n. 2, p. 191-209, 2009.

PAPINI-TERZI, F. S.; ROCHA, F. R.; VÊNCIO, R. Z. N.; FELIX, J. M.; BRANCO, D. S.; WACLAWOVSKY, A. J.; DEL BEM, L. E. V.; LEMBKE, C. G.; COSTA, M. D. L.; NISHIYAMA, M. Y.;

VICENTINI, R.; VINCENTZ, M. G. A.; ULIAN, E. C.; MENOSSI, M.. Sugarcane genes associated with sucrose content. **BMC genomics**, v. 10, p. 120, 2009.

PIPERIDIS, N.; JACKSON, P. A.; D'HONT, A.; BESSE, P.; HOARAU, J.-Y.; COURTOIS, B.; AITKEN, K. S.; MCINTYRE, C. L. Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. **Molecular Breeding**, v. 21, n. 2, p. 233-247, 2008.

RABOIN, L. –M.; PAUQUET, J.; BUTTERFIELD, M.; D'HONT, A.; GLASZMANN, J. –C.. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. **TAG. Theoretical and applied genetics**, v. 116, n. 5, p. 701-14, 2008.

RABOIN, L.-M.; OLIVEIRA, K. M.; LECUNFF, L.; TELISMART, H.; ROQUES, D.; BUTTERFIELD, M.; HOARAU, J.-Y.; D'HONT, A. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. **TAG - Theoretical and applied genetics.** v. 112, n. 7, p. 1382-1391, 2006.

RAFALSKI, A; TINGEY, S. SNPs and their use in maize. **Plant genotyping II: SNP TECHNOLOGY.** Oxfordshire, UK: CAB International, p. 30-43, 2008.

ROCHA, F. R.; PAPINI-TERZI, F. S.; NISHIYAMA, M. Y.; VÊNCIO, R. Z. N.; VICENTINI, R.; DUARTE, R. D. C.; DE ROSA, V. E.; VINAGRE, F.; BARSALOBRES, C.; MEDEIROS, A. H.; RODRIGUES, F. A.; ULIAN, E. C.; ZINGARETTI, S. M.; GALBIATTI, J. A.; ALMEIDA, R. S.; FIGUEIRA, A. V. O.; HEMERLY, A. S.; SILVA-FILHO, M. C.; MENOSSI, M.. Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. **BMC genomics**, v. 8, p. 71, 2007.

SULIMAN-POLLATSCHEK, S.; KASHKUSH, K.; SHATS, H.; HILLEL, J.; LAVI, U. Generation and mapping of AFLP, SSRs and SNPs in *Lycopersicon esculentum*. **Cellular and Molecular Biology Letters**, v. 7, n. 2A, p. 583-597, 2002.

TANAKA, T. [International HapMap project]. **Nippon rinsho. Japanese journal of clinical medicine**, v. 63 Suppl 1, p. 29-34, 2005.

TANG, J. VOSMAN, B. VOORRIPS, R. E. LINDEN, C. G. VAN DER; LEUNISSEN, J. A M. QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. **BMC bioinformatics**, v. 7, p. 438, 2006.

TENAILLON, M. I.; SAWKINS, M. C.; LONG, A. D.; GAUT, R. L.; DOEBLEY, J. F.; GAUT, B. S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). **Proceedings of the National Academy of Sciences of the United States of America.** v. 98, n. 16, p. 9161-9166, 2001.

WU, K. K.; BURNQUIST, W.; SORRELLS, M. E.; Tew, T. L.; Moore, P. H.; Tanksley, S. D. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theoretical and Applied Genetics**, v. 83, n. 3, p. 294-300, 1992.

YU, J.; HU, S.; WANG, J.; et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). **Science**, v. 296, p. 79-92, 2002.

CONCLUSÕES

– Foram desenvolvidos 856 locos EST-SSRs e 942 locos SNPs durante este trabalho. Esses marcadores moleculares funcionais representam um valioso recurso genético-genômico para utilização na análise da diversidade e mapeamento genético em cana-de-açúcar.

- Embora grande número de marcadores tenha sido desenvolvido, o banco de dados do SUCEST pode ainda ser explorado para o desenvolvimento de mais marcadores. A abundância de SSRs e SNPs encontrados nas regiões transcritas do genoma da cana-de-açúcar faz do banco de ESTs (SUCEST) uma valiosa fonte para o desenvolvimento de marcadores moleculares de maneira simples, rápida, econômica e muito eficiente.

- Os marcadores moleculares do tipo SNPs desenvolvidos para cana-de-açúcar quando avaliados com espectrometria de massas, possibilitam que a dosagem de cada loco SNP nos genótipos em estudo seja determinada com exatidão. Este resultado amplia os tipos de segregação alélica que podem ser utilizados no mapeamento genético de poliplóides como a cana-de-açúcar.

- Na condição específica do SNP estar em cópia única e ser polimórfico nos genitores, ele se comporta como um marcador codominante 1:2:1 em cana-de-açúcar. A adição de um novo tipo de segregação genética (1:2:1) ao mapeamento de cana-de-açúcar promoveu grande progresso. Todos os outros tipos de marcadores (SSR, RFLPs, RAPD, AFLP entre outros) até então utilizados para o mapeamento genético em cana-de-açúcar comportam-se como dominantes.

- Os marcadores moleculares desenvolvidos neste trabalho possibilitaram a construção de um mapa genético com maior densidade de marcas que o anterior existente para a população IACSP 95-3018 x IACSP 93-3046.

PERSPECTIVAS

Novas metodologias de análise genético-estatísticas precisam ser desenvolvidas para a análise genética dos marcadores SNPs presentes em cópias múltiplas em espécies poliplódes.

Com o desenvolvimento de novas metodologias de análise genético-estatísticas de SNPs pretende-se efetuar a integração de locos presentes em doses múltiplas no genoma da cana-de-açúcar, ao mapa-genético molecular construído neste trabalho. Um novo mapa genético-molecular para a população F_1 do cruzamento IACSP 95-3018 x IACSP 93-3046 será construído, no qual todos os locos EST-SSRs e SNPs genotipados na população F_1 e polimórficos entre os genitores, serão re-analisados utilizando-se as segregações alélicas identificadas pelos SNPs. O mapa resultante será muito melhor saturado, possibilitando a identificação de QTLs de diferentes efeitos com maior exatidão.

REFERÊNCIAS BIBLIOGRÁFICAS

AITKEN, K. S.; HERMANN, S.; KARNO, K.; BONNETT, G. D.; MCINTYRE, L. C.; JACKSON, P. A. Genetic control of yield related stalk traits in sugarcane. **TAG. Theoretical and applied genetics**, v. 117, n. 7, p. 1191-1203, 2008.

AITKEN, K. S.; JACKSON, P. A.; MCINTYRE, C. L. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. **TAG. Theoretical and applied genetics.**, v. 110, n. 5, p. 789-801, 2005.

AITKEN, K. S.; JACKSON, P. A.; MCINTYRE, C. L. Construction of a genetic linkage map for *Saccharum officinarum* incorporating both simplex and duplex markers to increase genome coverage. **Genome Research**, v. 50, n. 8, p. 742-756, 2007.

ALIX, K.; BAURENS, F. C.; PAULET, F.; GLASZMANN, J. C.; D'HONT, A. Isolation and characterization of a satellite DNA family in the *Saccharum* complex. **Genome**, v. 41, n. 6, p. 854-864, 1998.

ALIX, K.; PAULET, F.; GLASZMANN, J.-C.; D'HONT, A. Inter- Alu -like species-specific sequences in the *Saccharum* complex. **TAG Theoretical and Applied Genetics**, v. 99, n. 6, p. 962-968, 1999.

AL-JANABI, S. M.; MCCLELLAND, M.; PETERSEN, C.; SOBRAL, B. W. S. Phylogenetic analysis of organellar DNA sequences in the *Andropogoneae: Saccharinae*. **Theoretical and Applied Genetics**, v. 88, n. 8, p. 933-944, 1994.

AL-JANABI, S. M.; PARMESSUR, Y.; KROSS, H.; DHAYAN, S.; SAUMTALLY, S.; RAMDOYAL, K.; AUTREY, L. J. C.; DOOKUN-SAUMTALLY, A. Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. **Molecular Breeding**, v. 19, n. 1, p. 1-14, 2006.

BADGER, P. Ethanol from cellulose: A general review. **Trends in new crops and new uses. ASHS Press, Alexandria, VA**, p. 17-21, 2002.

BAIN, J. M.; TAVANTI, A.; DAVIDSON, A. D.; JACOBSEN, M. D.; SHAW, D.; GOW, N. A. R.; ODDS, F. C. Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. **Journal of clinical microbiology**, v. 45, n. 5, p. 1469-1477, 2007.

BANSAL, A.; VAN DEN BOOM, D.; KAMMERER, S.; HONISCH, C.; ADAM, G.; CANTOR, C. R.; KLEYN, P.; BRAUN, A. Association testing by DNA pooling: an effective initial screen. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 26, p. 16871-16874, 2002.

BARRENECHE, T.; BODENES, C.; LEXER, C.; TRONTIN, J.-F.; FLUCH, S.; STREIFF, R.; PLOMION, C.; ROUSSEL, G.; STEINKELLNER, H.; BURG, K.; FAVRE, J.-M.; GLÖSSL, J.; KREMER, A. A

genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. **TAG Theoretical and Applied Genetics**, v. 97, n. 7, p. 1090-1103, 1998.

BATLEY, J.; BARKER, G.; O´SULLIVAN, H.; EDWARDS, K. J.; EDWARDS, D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. **Plant physiology**, v. 132, n. 1, p. 84-91, 2003.

BEAULIEU, M.; KOSMAN, D. Molecular haplotyping combining a novel AS-PCR technique with base-specific cleavage and mass spectrometry. SNPs, Haplotypes and Cancer: Application in Molecular Epidemiology. Conference Proceedings, Key Biscayne, Florida, September 13-17, A60. **Anais**, 2003.

BELL, C. J.; ECKER, J. R. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. **Genomics**, v. 19, n. 1, p. 137–144, 1994.

BESSE, P. MCINTYRE, C. L.; BERDING, N. Ribosomal DNA variations in *Erianthus*, a wild sugarcane relative (*Andropogoneae-Saccharinae*). **Theoretical and Applied Genetics**, v. 92, n. 6, p. 733-743, 1996.

BIDOIA, M. Instalação, condução e colheita de experimentos. **Cana-de-açúcar. Campinas: IAC**. p. 821-838, 2008.

BOYLE, R.; GREENWOOD, C.; HOHLER, A.; LIEBREICH, M.; SONNTAG-O'BRIEN, V.; TYNE, A.; USHER, E. Global trends in sustainable energy investment. **United Nations Foundation**, 2008.

BRANDES, E. Origin, dispersal, and use in breeding of the Melanesian garden sugarcanes and their derivatives, *Saccharum officinarum* L. Proceedings of the International Society of Sugar Cane Technologists. **Anais,** v. 9, p. 709–750, 1956.

BRANDES, E.; SARTORIS, G.; GRASSL, C. Assembling and evaluating wild forms of sugarcane and closely related plants. Proc. ISSCT. **Anais,** v. 6, p. 128–151, 1938.

BREMER, G. Problems in breeding and cytology of sugar cane. **Euphytica**, v. 10, n. 1, p. 59–78, 1961.

BREMER, G. The cytology of *Saccharum*. **Proc. Congr. Intern. Soc. Sugar Cane Techn.**, 1930.

BREMER, G. The cytology of sugarcane. A cytological investigation of some cultivated kinds and their parents. **Genetica**, v. 6, p. 497–525, 1924.

BRESSIANI, J. A.; VENCOVSKY, R.; BURNQUIST, W.L. Interação entre famílias de cana-de-açúcar e locais: efeito na resposta esperada com a seleção. **Bragantia**, v. 61, n. 1, p. 01–10, 2002.

BUETOW, K. H.; EDMONSON, M. N.; CASSIDY, A. B. Reliable identification of large numbers of candidate SNPs from public EST data. **Nature Genetics**, v. 21, n. 3, p. 323-325, 1999.

BULL, J.; HOGARTH, D.; BASFORD, K. Impact of genotype multiply environment interactions on response to selection in sugarcane. **Australian Journal of Experimental Agriculture, Melbourne**, v. 32, n. 6, p. 731–737, 1992.

BUNDOCK, P. C.; ELIOTT, F. G.; ABLETT, G.; BENSON, A. D.; CASU, R. E.; AITKEN, K. S.; HENRY, R. J. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. **Plant biotechnology journal**, v. 7, n. 4, p. 347-354, 2009.

BUNDOCK, P. C.; HENRY, R. J. Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. **TAG. Theoretical and applied genetics.**, v. 109, n. 3, p. 543-551, 2004.

BURNER, D. Cytogenetic analyses of sugarcane relatives (*Andropogoneae: Saccharinae*). **Euphytica**, v. 54, n. 1, p. 125–133, 1991.

BURNQUIST, W.L.; SORRELLES, M.; TANKSLEY, S. Characterization of genetic variability in Saccharum germplasm by means of restriction fragment length polymorphism (RFLP) analysis. Proc. Int. Soc. Sugar Cane Technol. **Anais**, v. 21, n. 2, p. 355-365, 1995.

BURTON, P. R.; CLAYTON, D. G.; CARDON, L. R.; et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. **Nature genetics**, v. 39, n. 11, p. 1329-1337, 2007.

BUTCHER, A.; WILLIAMS, R.; WHITAKER, D.; LING, S.; SPEED, P.; MORAN, F. Improving linkage analysis in outcrossed forest trees - an example from *Acacia mangium*. **TAG. Theoretical and applied genetics**, v. 104, n. 6-7, p. 1185-1191, 2002.

CAMARGO, L. Mapeamento de QRL's: contribuições e perspectivas. Jornada em Genética e Melhoramento de Plantas, Goiânia, GO. Análise de QTL no Melhoramento de Plantas. Goiânia: FUNAPE. **Anais,** p. 49–61, 2000.

CARLIER, J. D.; REIS, A.; DUVAL, M. F.; COPPENS D'EECKENBRUGGE, G.; LEITAO, J. M. Genetic maps of RAPD, AFLP and ISSR markers in *Ananas bracteatus* and *A. comosus* using the pseudo-testcross strategy. **Plant Breeding**, v. 123, n. 2, p. 186-192, 2004.

CHO, R. J.; MINDRINOS, M.; RICHARDS, D. R.; SAPOLSKY, R. J.; ANDERSON, M.; DRENKARD, E.; DEWDNEY, J.; REUBER, T. L.; STAMMERS, M.; FEDERSPIEL, N.; THEOLOGIS, A.; YANG, W. H.; HUBBELL, E.; AU, M., CHUNG, E. Y.; LASHKARI, D.; LEMIEUX, B.; DEAN, C.; LIPSHUTZ, R. J.; AUSUBEL, F. M.; DAVIS, R. W.; OEFNER, P. J. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. **Nature genetics**, v. 23, n. 2, p. 203-207, 1999.

CONAB, 2011.
http://www.conab.gov.br/OlalaCMS/uploads/arquivos/11_05_27_11_53_13_boletim_cana_portugues_-_maio_2011_1o_lev..pdf

CORDEIRO, G. M.; CASU, R.; MCINTYRE, C. L.; MANNERS, J. M.; HENRY, R. J. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *erianthus* and *sorghum*. **Plant science**, v. 160, n. 6, p. 1115-1123, 2001.

CORDEIRO, G. M.; ELIOTT, F.; MCINTYRE, C. L.; CASU, R. E.; HENRY, R. J. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. **TAG. Theoretical and applied genetics,** v. 113, n. 2, p. 331-343, 2006.

CORDEIRO, G. M.; TAYLOR, G.; HENRY, R. J. Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. **Plant science**, v. 155, n. 2, p. 161–168, 2000.

D'HONT, A.; GRIVET, L.; FELDMANN, P.; RAO, S.; BERDING, N.; GLASZMANN, J. C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Molecular & general genetics : MGG**, v. 250, n. 4, p. 405-413, 1996.

D'HONT, A.; LU, Y.; FELDMANN, P.; GLASZMANN, J. Cytoplasmic diversity in sugar cane revealed by heterologous probes. Proceedings - West Indies sugar technologists twenty-fifth conference, Belize, April 18-21, 1994, The Sugar Association of the Caribbean (SAC). **Anais**, 1993.

D'HONT, A.; RAO, P.; FELDMANN, P.; GRIVET, L.; ISLAM-FARIDI, N.; BERDING, N.; GLASZMANN, J. C. Identification and characterization of intergeneric hybrids, *Saccharum officinarum* x *Erianthus arundinaceus*, with molecular markers and in situ hybridization. **Theoretical and Applied Genetics**, v. 91, p. 320–326, 1995.

D'HONT, A.; SOUZA, G. M.; MENOSSI, M.; VINCENTZ, M.; VAN-SLUYS, M. A.; GLASZMANN, J. C.; ULIAN, E. Sugarcane: a major source of sweetness, alcohol, and bio-energy. **Genomics of tropical crop plants**, p. 483-513, 2008.

DA SILVA, J. A. G.; HONEYCUTT, R. J.; BURNQUIST, W.; AL-JANABI, S. M.; SORRELLS, M. E.; TANKSLEY, S. D.; SOBRAL, B. W. S. *Saccharum spontaneum* L. "SES 208" genetic linkage map combining RFLP- and PCR-based markers. **Molecular Breeding**, v. 1, n. 2, p. 165-179, 1995.

DA SILVA, J. A. G.; SORRELLS, M. E.; BURNQUIST, W. L.; TANKSLEY, S. D. RFLP linkage map and genome analysis of *Saccharum spontaneum*. **Genome**, v. 36, n. 4, p. 782-791, 1993.

DANIELS, J.; ROACH, B. Taxonomy and evolution in sugarcane. **Sugarcane improvement through breeding.** Amsterdam: Elsevier, p. 7–84, 1987.

DAUGROIS, J. H.; GRIVET, L.; ROQUES, D.; HOARAU, J. Y.; LOMBARD, H.; GLASZMANN, J. C.; D'HONT, A. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar "R570." **Theoretical and Applied Genetics**, v. 92, n. 8, p. 1059-1064, 1996.

DECROOCQ, V.; FAVÉ, M. G.; HAGEN, L.; BORDENAVE, L.; DECROOCQ, S. Development and transferability of apricot and grape EST microsatellite markers across taxa. **TAG. Theoretical and applied genetics**, v. 106, n. 5, p. 912-922, 2003.

DEMIRBAS, A. Biomass resource facilities and biomass conversion processing for fuels and chemicals. **Energy Conversion and Management**, v. 42, n. 11, p. 1357-1378, 2001.

DOWNES, K.; BARRATT, B. J.; AKAN, P.; BUMPSTEAD, S. J.; TAYLOR, S. D.; CLAYTON, D. G. SNP allele frequency estimation in DNA pools and variance components analysis. **BioTechniques**, v. 36, n. 5, p. 840-845, 2004.

EDWARDS, K J; BARKER, J. H. DALY, A. JONES, C.; KARP, A. Microsatellite libraries enriched for several microsatellite sequences in plants. **BioTechniques**, v. 20, n. 5, p. 758-760, 1996.

EHRICH, M.; NELSON, M. R.; STANSSENS, P.; ZABEAU, M.; LILOGLOU, T.; XINARIANOS, G.; CANTOR, C. R.; FIELD, J. K.; VAN DEN BOOM, D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 44, p. 15785-15790, 2005.

ENGLISH, B.C.; MENARD, J.R.; UGARTE, D.T. E WALSH, M. Economic impacts of ethanol production from maize stover in selected midwestern states. In: Outlaw, J. **Agriculture as a Producer and Consumer of Energy**. Knoxville: CABI Publishing, 2005

EUJAYL, I.; SLEDGE, M. K.; WANG, L.; MAY, G. D.; CHEKHOVSKIY, K.; ZWONITZER, J. C.; MIAN, M. A. R. *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. **TAG. Theoretical and applied genetics.**, v. 108, n. 3, p. 414-422, 2004.

FAO, FAOSTAT – Food and Agricultural Organization of the United Nations FAO Statitics Division. http://faostat.fao.org/. 2009

FAPESP. Genoma. **São Paulo**, v. 1, 2001.

FEDERER, W. T. Augmented (or hoonuiaku) designs. **Hawaiian Planters Record**, v. 55, n. 2, p. 191–208, 1956.

FERGUSON, B.; STREET, S. L.; WRIGHT, H.; PEARSON, C.; JIA, Y.; THOMPSON, S. L.; ALLIBONE, P.; DUBAY, C. J.; SPINDEL, E.; NORGREN, R. B. Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). **BMC genomics**, v. 8, p. 43, 2007.

FRANCIA, E.; TACCONI, G.; CROSATTI, C.; BARABASCHI, D.; BULGARELLI, D.; DALL'AGLIO, E.; VALÈ, G. Marker assisted selection in crop plants. **Plant Cell Tissue and Organ Culture**, v. 82, n. 3, p. 317-342, 2005.

FRAZER, K. A; BALLINGER, D. G.; COX, D. R.; et al. A second generation human haplotype map of over 3.1 million SNPs. **Nature**, v. 449, n. 7164, p. 851-861, 2007.

GARCIA, A. A. F.; KIDO, E. A; MEZA, A. N.; SOUZA, H. M. B.; PINTO, L. R.; PASTINA, M. M.; LEITE, C. S.; DA SILVA, J.; FIGUEIRA, A.; ULIAN, E. C.; SOUZA, A. P. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **TAG. Theoretical and applied genetics**, v. 112, n. 2, p. 298-314, 2006.

GIANCOLA, S.; MCKHANN, H. I.; BÉRARD, A.; CAMILLERI, C.; DURAND, S.; LIBEAU, P.; ROUX, F.; REBOUD, X.; GUT, I. G.; BRUNEL, D. Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. **TAG. Theoretical and applied genetics.**, v. 112, n. 6, p. 1115-1124, 2006.

GIBBS, R. A.; BELMONT, J.; HARDENBOL, P. The International HapMap Project. **Nature**, v. 426, n. 6968, p. 789-796, 2003.

GLASZMANN, J. C.; LU, Y. H.; LANAUD, C. Variation of nuclear ribosomal DNA in sugarcane. **Journal of Genetics & Breeding**, n. 44, p. 191-198, 1990.

GOLDEMBERG, J. Ethanol for a sustainable energy future. **Science**, v. 315, n. 5813, p. 808-810, 2007.

GRABER, J. H.; SMITH, C. L.; CANTOR, C. R. Differential sequencing with mass spectrometry. **Genetic analysis : biomolecular engineering**, v. 14, n. 5-6, p. 215-219, 1999.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic Linkage Maps of *Eucalyptus grandis* and *Eucalyptus urophylla* Using a Pseudo-Testcross: Mapping Strategy and RAPD Markers. **Genetics**, v. 137, n. 4, p. 1121-1137, 1994.

GRIFFIN, T. J.; SMITH, L. M. Single-nucleotide polymorphism analysis by MALDI–TOF mass spectrometry. **Trends in Biotechnology**, v. 18, n. 2, p. 77-84, 2000.

GRIFFITHS, A.; BARRETT, B.; ELLISON, N.; SAWBRIDGE, T.; SPANGENBURG, G.; BRYAN, G. J.; SCHREIBER, M. Getting out there: transportability of white clover EST and genomic DNA-derived microsatellites. Proceedings of the 12th Australasian plant breeding conference, Perth. **Anais,** p. 15–20, 2002.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current opinion in plant biology**, v. 5, n. 2, p. 122-127, 2002.

GRIVET, L.; D'HONT, A.; ROQUES, D.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J. C. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. **Genetics**, v. 142, n. 3, p. 987-1000, 1996.

GRIVET, L.; DANIELS, C.; GLASZMANN, J. C.; D'HONT, A. A review of recent molecular genetics evidence for sugarcane evolution and domestication. **Ethnobotany Research & Applications**, v. 2, p. 9-17, 2004.

GRIVET, L.; GLASZMANN, J. J. C.; ARRUDA, P.; GLASZMANN, C. Sequence polymorphism from EST data in sugarcane: a fine analysis of 6-phosphogluconate dehydrogenase genes. **Genetics and molecular biology**, v. 24, n. 1-4, p. 161–167, 2001.

GRIVET, L.; GLASZMANN, J-C.; VINCENTZ, M.; DA SILVA, F.; ARRUDA, P. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. **TAG. Theoretical and applied genetics**, v. 106, n. 2, p. 190-197, 2003.

GUIMARÃES, C. T. **Mapeamento comparativo e detecção de QTLs em cana-de-açúcar utilizando marcadores moleculares**, Viçosa, MG: UFV, 1999. 70p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 1999.

GUPTA, P. K.; ROY, J. K.; PRASAD, M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, v. 80, n. 4, p. 524–535, 2001.

GUT, I. G. DNA analysis by MALDI-TOF mass spectrometry. **Human mutation**, v. 23, n. 5, p. 437-441, 2004.

HARVEY, M.; D'HONT, A.; ALIX, K.; HUCKETT, B. Use of PCR based markers for identification of Erianthus genetic material in putative intergeneric hybrids (*Saccharum* x *Erianthus*). Proc. S. Afr. Sug. Technol. Ass. **Anais,** v. 72, p.72: 218–320, 1998.

HEINZ, D.; TEW, T. Hibridization procedures. In: **Sugarcane improvement through breeding**. Amsterdam, ND: Elsevier, p. 313-342, 1987.

HILL, W. G.; GODDARD, M. E.; VISSCHER, P. M. Data and theory point to mainly additive genetic variance for complex traits. **PLoS genetics**, v. 4, n. 2, p. e1000008, 2008.

HOARAU, J.-Y.; OFFMANN, B.; D'HONT, A.; ROQUES, D.; RISTERUCCI, A.-M.; GLASZMANN, J.-C.; GRIVET, L. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. **TAG Theoretical and Applied Genetics**, v. 103, n. 1, p. 84-97, 2001.

HOGARTH, D. M.; BULL, J. K. The implications of genotype x environment interactions for evaluating sugarcane families. I. Effect on selection. **In: KANG, M.S. GE interaction and plant breeding.** Baton Rouge: Louisiana State University, p.335-346, 1990.

HONISCH, C.; RAGHUNATHAN, A.; CANTOR, C. R.; PALSSON, B. Ø.; VAN DEN BOOM, D. High-throughput mutation detection underlying adaptive evolution of *Escherichia coli-K12*. **Genome research**, v. 14, n. 12, p. 2495-2502, 2004.

IRWIN, D. The MASSARRAY System for Plant Genomics. In: R.J. Henry (Ed.); **Plant Genotyping II SNP Technology.** Oxfordshire, UK:CABI Publishing. 285 p, 2008.

JACKSON, P.; MCRAE, T.; HOGARTH, M. Selection of sugarcane families across variable environments I. Sources of variation and an optimal selection index. **Field Crops Research**, v. 43, n. 2-3, p. 109–118, 1995a.

JACKSON, P.; MCRAE, T.; HOGARTH, M. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. **Field Crops Research**, v. 43, n. 2-3, p. 119–130, 1995b.

JANNOO, N.; GRIVET, L.; DOOKUN, A.; D'HONT, A.; GLASZMANN, J. Linkage disequilibrium among modern sugarcane cultivars. **TAG Theoretical and Applied Genetics**, v. 99, n. 6, p. 1053–1060, 1999.

JIANG, C.; ZENG, Z.-B. Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. **Genetica**, v. 101, n. 1, p. 47-58, 1997.

JURINKE, C.; DENISSENKO, M. F.; OETH, P.; EHRICH, M.; VAN DEN BOOM, D.; CANTOR, C. R. A single nucleotide polymorphism based approach for the identification and characterization of gene expression modulation using MassARRAY. **Mutation research**, v. 573, n. 1-2, p. 83-95, 2005.

KAMMERER, S.; ROTH, R. B.; HOYAL, C. R.; RENELAND, R.; MARNELLOS, G.; KIECHLE, M.; SCHWARZ-BOEGER, U.; GRIFFITHS, L. R.; EBNER, F.; REHBOCK, J.; CANTOR, C. R.; NELSON, M. R.; BRAUN, A. Association of the NuMA region on chromosome 11q13 with breast cancer susceptibility. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 6, p. 2004-2009, 2005.

KAMMERER, S.; ROTH, R. B.; RENELAND, R.; MARNELLOS, G.; HOYAL, C. R.; MARKWARD, N. J.; EBNER, F.; KIECHLE, M.; SCHWARZ-BOEGER, U.; GRIFFITHS, L. R.; ULBRICH, C.; CHROBOK, K.; FORSTER, G.; PRAETORIUS, G. M.; MEYER, P.; REHBOCK, J.; CANTOR, C. R.; NELSON, M. R.; BRAUN, A. Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus. **Cancer research**, v. 64, n. 24, p. 8906-8910, 2004.

KANAZIN, V.; TALBERT, H.; SEE, D.; DECAMP, P.; NEVO, E.; BLAKE, T. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). **Plant molecular biology**, v. 48, n. 5-6, p. 529-537, 2002.

KARAS, M.; HILLENKAMP, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. **Analytical chemistry**, v. 60, n. 20, p. 2299-2301, 1988.

KIM, S.; MISRA, A. SNP genotyping: technologies and biomedical applications. **Annual review of biomedical engineering**, v. 9, p. 289-320, 2007.

KÖSTER, H. Massarray™: Highly Accurate and Versatile High Throughput Analysis of Genetic Variations. **Mass Spectrometry and Genomic Analysis**. p.33–49, 2002.

KOTA, R.; RUDD, S.; FACIUS, A.; KOLESOV, G.; THIEL, T.; ZHANG, H.; STEIN, N.; MAYER, K.; GRANER, A. Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). **Molecular genetics and genomics : MGG**, v. 270, n. 1, p. 24-33, 2003.

KUMAR, L. S. DNA markers in plant improvement: An overview. **Plant Molecular Biology**, v. 17, p. 143-182, 1999.

LANDELL, M. G. A.; BRESSIANI, J. A. **Melhoramento genético, caracterização e manejo varietal**. p. 101-155, 2008.

LANDER, E. S.; GREEN, P. Construction of multilocus genetic linkage maps in humans. **Proceedings of the National Academy of Sciences of the United States of America**, v. 84, n. 8, p. 2363-2367, 1987.

LI, H. W.; PRINCE, S. Chromosome numbers of noble sugarcane clones. Proceding. Int. Soc. of Sugarcane Technology, **Anais**, n. 12, p. 884-886, 1967.

LI, Y.-C.; KOROL, A. B.; FAHIMA, T.; BEILES, A.; NEVO, E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. **Molecular ecology**, v. 11, n. 12, p. 2453-2465, 2002.

LIMA, M. L. A.; GARCIA, A. A. F.; OLIVEIRA, K. M.; MATSUOKA, S.; ARIZONO, H.; DE SOUZA, C. L.; DE SOUZA, A. P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). **TAG Theoretical and Applied Genetics**, v. 104, n. 1, p. 30-38, 2002.

LIN, M.; LOU, X.; CHANG, M.; WU, R. A General Statistical Framework for Mapping Quantitative Trait Loci in Nonmodel Systems: Issue for Characterizing Linkage Phases. **Genetics**, v. 165, p. 901-913, 2003.

LU, Y. H.; D'HONT, A.; PAULET, F.; GRIVET, L.; ARNAUD, M.; GLASZMANN, J. C. Molecular diversity and genome structure in modern sugarcane varieties. **Euphytica**, v. 78, n. 3, p. 217-226, 1994b.

LU, Y.; D'HONT, A.; WALKER, D.; RAO, P. S.; FELDMANN, P.; GLASZMANN, J. C. Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. **Euphytica**, v. 78, n. 1, p. 7–18, 1994a.

MACHADO, G.; DA SILVA, W. M.; IRVINE, J. E. Sugar cane breeding in Brazil: the Copersucar Program. **Copersucar International Sugarcane Breeding Workshop,** Cooperativa de

Produtores de Cana, Açúcar e Álcool do Estado de São Paulo Ltda, São Paulo, SP, Brazil. p. 215–232, 1987.

MALIEPAARD, C.; ALSTON, F. H.; VAN ARKEL, G.; BROWN, L. M.; CHEVREAU, E.; DUNEMANN, F.; EVANS, K. M.; GARDINER, S.; GUILFORD, P.; VAN HEUSDEN, A. W.; JANSE, J.; LAURENS, F.; LYNN, J. R.; MANGANARIS, A. G.; DEN NIJS, A. P. M.; PERIAM, N.; RIKKERINK, E.; ROCHE, P.; RYDER, C.; SANSAVINI, S.; SCHMIDT, H.; TARTARINI, S.; VERHAEGH, J. J.; VRIELINK-VAN GINKEL, M.; KING, G. J. Aligning male and female linkage maps of apple (*Malus pumila Mill.*) using multi-allelic markers. **TAG Theoretical and Applied Genetics**, v. 97, n. 1-2, p. 60-73, 1998.

MALIEPAARD, C.; JANSEN, J.; VAN OOIJEN, J. W. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetical Research**, v. 70, n. 3, p. 237-250, 1997.

MANGELSDORF, A. J. Sugarcane breeding in Hawaii. Part II – 1921-1952. **Hawaii. Plant. Rec**, n. 54, p. 101-137, 1953.

MANNERS, J.; MCINTYRE, L.; CASU, R.; CORDEIRO, G.; JACKSON, M.; AITKEN, K.; JACKSON, P.; BONNETT, G.; LEE, S.; HENRY, R. Can genomics revolutionize genetics and breeding in sugarcane. Proceedings of the 4th International Crop Science Congress. Brisbane, Australia. **Anais,** v. 26, 2004.

MARGARIDO, G. R. A.; SOUZA, A. P.; GARCIA, A. A. F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, v. 144, n. 3, p. 78-79, 2007.

MASOULEH, A. K.; WATERS, D. L. E.; REINKE, R. F.; HENRY, R. J. A high-throughput assay for rapid and simultaneous analysis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI-TOF mass spectrometry. **Plant biotechnology journal**, v. 7, n. 4, p. 355-363, 2009.

MATSUOKA, S.; GARCIA, A. A. F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: A. Borém (Ed.); **Melhoramento de espécies cultivadas. (2 ed).** Viçosa: UFV, p.205-251, 1999b.

MATSUOKA, S.; GARCIA, A. A. F.; CALHEIROS, G. G. Hibridação em Cana-de-açúcar. In: A. Borém (Ed.); **Hibridação Artificial de Plantas**. Viçosa: UFV, p.221-254, 1999a.

MCINTYRE, C.L.; JACKSON, M.; CORDEIRO, G. M.; AMOUYAL, O.; HERMANN, S.; AITKEN, K. S.; ELIOTT, F.; HENRY, R. J.; CASU, R. E.; BONNETT, G. D. The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. **Molecular Breeding**, v. 18, n. 1, p. 39-50, 2006.

MING, R.; MOORE, P. H.; WU, K.-K.; D'HONT, A.; GLASZMANN, J. C.; TEW, T. L.; MIRKOV, T. E.; DA SILVA, J.; JIFON, J.; RAI, M.; SCHNELL, R. J.; BRUMBLEY, S. M.; LAKSHMANAN, P.; COMSTOCK, J. C.; PATERSON, A. H. Sugarcane Improvement through Breeding and

Biotechnology. In: J. Janick (Ed.); **Plant Breeding Reviews Vol 27**, Oxford, UK: John Wiley & Sons, Inc., 2006.

MOGG, R.; BATLEY, J.; HANLEY, S.; EDWARDS, D.; O'SULLIVAN, H.; EDWARDS, J. Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. **TAG. Theoretical and applied genetics.**, v. 105, n. 4, p. 532-543, 2002.

MUKHERJEE, S. K. Origin and distribution of *Saccharum*. **Botanical Gazette**, v. 119, n. 1, p. 55–61, 1957.

NAIDU, K. M.; SREENIVASAN, T. Conservation of sugarcane germplasm. Proceedings of the Copersucar International Sugarcane Breeding Workshop. Copersucar Technology Centre, Piracicaba-SP, Brazil. **Anais,** p.33–53, 1987.

NAIR, N. V.; NAIR, S.; SREENIVASAN, T.; MOHAN, M. Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. **Genetic Resources and Crop Evolution**, v. 46, n. 1, p. 73–79, 1999.

OBERACHER, H. On the use of different mass spectrometric techniques for characterization of sequence variability in genomic DNA. **Analytical and Bioanalytical Chemistry**, v. 391, n. 1, p. 135-149, 2008.

OLIVEIRA, K. M.; PINTO, L. R.; MARCONI, T. G.; MARGARIDO, G. R. A.; PASTINA, M. M.; TEIXEIRA, L. H. M.; FIGUEIRA, A. V.; ULIAN, E. C.; GARCIA, A. A. F.; SOUZA, A. P. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. **Molecular Breeding**, v. 20, n. 3, p. 189-208, 2007.

OLIVEIRA, K. M.; PINTO, L. R.; MARCONI, T. G.; MOLLINARI, M.; ULIAN, E. C.; CHABREGAS, S. M.; FALCO, M. C.; BURNQUIST, W.; GARCIA, A. A. F.; SOUZA, A. P. Characterization of new polymorphic functional markers for sugarcane. **Genome Research**, v. 52, n. 2, p. 191-209, 2009.

ORTOLAN, M. Perspectivas para o setor sucroalcooleiro. **Tópicos em Tecnologia Sucroalcooleira.** Jaboticabal: Gráfica Multipress Ltda, p. 07–16, 2006.

PANJE, R. R.; BABU, C. N. Studies in *Saccharum spontaneum* distribution and geographical association of chromosome numbers. **Cytologia**, , n. 25, p. 150-152, 1960.

PETREN, K.; GRANT, B. R.; GRANT, P. R. A phylogeny of Darwin's finches based on microsatellite DNA length variation. **Proceedings of the Royal Society Biological Sciences**, v. 266, n. 1417, p. 321-329, 1999.

PINTO, L. R.; OLIVEIRA, K. M.; MARCONI, T. G.; GARCIA, A. A. F.; ULIAN, E. C.; SOUZA, A. P. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. **Plant Breeding**, v. 125, n. 4, p. 378-384, 2006.

PINTO, L. R.; OLIVEIRA, K. M.; ULIAN, E. C.; GARCIA, A. A. F.; SOUZA, A. P. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. **Genome**, v. 804, p. 795-804, 2004.

PIPERIDIS, G.; D'HONT, A. Chromosome composition analysis of various *Saccharum* interespecific hybrids by genomic in situ hybridization (GISH). Proc. Int. Soc. Technol. **Anais,** 24 p. 565-566, 2001.

PIPERIDIS, N.; JACKSON, P. A.; D'HONT, A.; BESSE, P.; HOARAU, J.-Y.; COURTOIS, B.; AITKEN, K. S.; MCINTYRE, C. L. Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait associations. **Molecular Breeding**, v. 21, n. 2, p. 233-247, 2008.

PORCEDDU, A.; ALBERTINI, E.; BARCACCIA, G.; FALISTOCCO, E.; FALCINELLI, M. Linkage mapping in apomictic and sexual Kentucky bluegrass (*Poa pratensis* L.) genotypes using a two way pseudo-testcross strategy based on AFLP and SAMPL markers. **TAG. Theoretical and applied genetics**, v. 104, n. 2-3, p. 273-280, 2002.

POWELL, W.; MACHRAY, G. C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in plant science**, v. 1, n. 7, p. 215–222, 1996.

PRICE, S. Cytological studies in *Saccharum* and allied genera. III. Chromosome numbers in interspecific hybrids. **Botanical Gazette**, v. 118, n. 3, p. 146–159, 1957.

PRICE, S. Cytology of *Saccharum robustum* and related symapatric species and natural hybrids. **Techical. Bulletin No. 1337**, U.S. Department of Agriculture, 1965.

PRICE, S.; DANIELS, J. Cytology of South Pacific sugarcane and related grasses, with special reference to Fiji. **Journal Heredity**, n. 59, p. 141-145, 1968.

RABOIN, L.-M.; OLIVEIRA, K. M.; LECUNFF, L.; TELISMART, H.; ROQUES, D.; BUTTERFIELD, M.; HOARAU, J.-Y.; D'HONT, A. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. **TAG - Theoretical and applied genetics.** v. 112, n. 7, p. 1382-1391, 2006.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current opinion in plant biology**, v. 5, n. 2, p. 94-100, 2002.

RAFALSKI, A; TINGEY, S. SNPs and their use in maize. **Plant genotyping II: SNP TECHNOLOGY.** Oxfordshire, UK: CAB International, p. 30-43, 2008.

RIDOUT, M. S.; TONG, S.; VOWDEN, C. J.; TOBUTT, K. R. Three-point linkage analysis in crosses of allogamous plant species. **Genetical Research**, v. 72, n. 2, p. 111-121, 1998.

RIPOL, M. I.; CHURCHILL, G. A.; DA SILVA, J. A. G.; SORRELLS, M. Statistical aspects of genetic mapping in autopolyploids. **Gene**, v. 235, p. 31-41, 1999.

RITTER, E.; SALAMINI, F. The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. **Genetical Research**, v. 67, n. 01, p. 55-65, 1996.

ROACH, B. Nobilisation of sugarcane. Proceedings of the International Society for Sugar Cane Technology. **Anais,** p. 14, p. 206–216, 1972.

ROSS, P.; HALL, L.; SMIRNOV, I.; HAFF, L. High level multiplex genotyping by MALDI-TOF mass spectrometry. **Nature biotechnology**, v. 16, n. 13, p. 1347-1351, 1998.

RUSSELL, J.; FULLER, J.; YOUNG, G.; THOMAS, B.; TARAMINO, G.; MACAULAY, M.; WAUGH, R.; POWELL, W. Discriminating between barley genotypes using microsatellite markers. **Genome research**, v. 40, n. 4, p. 442-450, 1997.

SENIOR, M. L. CHIN, E. C. L. LEE, M. SMITH, J. S. C.; STUBER, C. W. Simple sequence repeat markers developed from maize sequences found in the GENBANK database : Map construction. **Crop science**, v. 36, n. 6, p. 1676-1683, 1996.

SHEPHERD, M.; CROSS, M.; DIETERS, M. J.; HENRY, R. Genetic maps for *Pinus elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* using AFLP and microsatellite markers. **TAG. Theoretical and applied genetics.**, v. 106, n. 8, p. 1409-1419, 2003.

SKINNER, J.; HOGARTH, D.; WU, K. K. Sugarcane Improvement through Breeding. **Developments in Crop Science,** Amsterdam**:** Elsevier, v. 11, p. 409-453, 1987.

SOBRAL, B.; BRAGA, D.; LAHOOD, E.; KEIM, P. Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. subtribe of the Andropogoneae Dumort. tribe. **Theoretical and Applied Genetics**, v. 87, n. 7, p. 843–853, 1994.

SREENIVASAN, T. V.; AHLOOWALIA, B. S.; HEINZ, D. J. Cytogenetics. In Sugarcane improvement through breeding. **Edited by D.J. Heinz. Elsevier Press, Amsterdam**, p. 211-253, 1987.

STANSSENS, P.; ZABEAU, M.; MEERSSEMAN, G.; REMES, G.; GANSEMANS, Y.; STORM, N.; HARTMER, R.; HONISCH, C.; RODI, C. P.; BÖCKER, S.; VAN DEN BOOM, D. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. **Genome research**, v. 14, n. 1, p. 126-133, 2004.

STERKY, F.; LUNDEBERG, J. Sequence analysis of genes and genomes. **Journal of biotechnology**, v. 76, n. 1, p. 1-31, 2000.

STEVENSON, G. C. **Genetics and breeding of sugarcane**. London: Longmans, 284p, 1965.

SULIMAN-POLLATSCHEK, S.; KASHKUSH, K.; SHATS, H.; HILLEL, J.; LAVI, U. Generation and mapping of AFLP, SSRs and SNPs in *Lycopersicon esculentum*. **Cellular and Molecular Biology Letters**, v. 7, n. 2A, p. 583-597, 2002.

SYBENGA, J. **Cytogenetics in plant breeding.** Springer-Verlag, 469p, 1992.

TANG, K.; OETH, P.; KAMMERER, S.; DENISSENKO, M. L. F.; EKBLOM, J.; JURINKE, C.; VAN DEN BOOM, D.; BRAUN, A.; CANTOR, C. R. Mining Disease Susceptibility Genes through SNP Analyses and Expression Profiling Using MALDI-TOF Mass Spectrometry. **Journal of Proteome Research**, v. 3, n. 2, p. 218-227, 2004.

TENAILLON, M. I.; SAWKINS, M. C.; LONG, A. D.; GAUT, R. L.; DOEBLEY, J. F.; GAUT, B. S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). **Proceedings of the National Academy of Sciences of the United States of America.** v. 98, n. 16, p. 9161-9166, 2001.

TOST, J.; GUT, I. G. Genotyping single nucleotide polymorphisms by mass spectrometry. **Mass spectrometry reviews**, v. 21, n. 6, p. 388-418, 2003.

UMEMOTO, T.; AOKI, N.; LIN, H.; NAKAMURA, Y.; INOUCHI, N.; SATO, Y.; YANO, M.; HIRABAYASHI, H.; MARUYAMA, S. Natural variation in rice starch synthase IIa affects enzyme and starch properties. **Functional plant biology**, v. 31, n. 7, p. 671–684, 2004.

UNICA - União da Indústria de Cana-de-açúcar. Setor Sucroalcooleiro – Histórico. Disponível em: http://www.unica.com.br. Acesso em: 06 jul. 2008

UNICA - União da Indústria de Cana-de-açúcar. Setor Sucroalcooleiro – Histórico. Disponível em: http://www.unica.com.br. Acesso em: 20 mai. 2011.

VARSHNEY, R. K.; BEIER, U.; KHLESTKINA, E. K.; KOTA, R.; GRANER, A.; KORZUN, V.; BÖRNER, A. Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency, and applications for genome mapping and diversity studies. **TAG. Theoretical and applied genetics.** v. 114, n. 6, p. 1105-1116, 2007.

VETTORE, A. L.; DA SILVA, F. R.; KEMPER, E. L.; et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome research**, v. 13, n. 12, p. 2725-2735, 2003.

VETTORE, A. L.; SILVA, F. R.; KEMPER, E. L.; ARRUDA, P. The libraries that made SUCEST. **Genetics and Molecular Biology**, v. 24, p. 1-7, 2001.

WU, K. K.; BURNQUIST, W.; SORRELLS, M. E.; Tew, T. L.; Moore, P. H.; Tanksley, S. D. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theoretical and Applied Genetics**, v. 83, n. 3, p. 294-300, 1992.

WU, R.; MA, C.-X.; PAINTER, I.; ZENG, Z.-B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. **Theoretical population biology**, v. 61, n. 3, p. 349-363, 2002a.

WU, R.; MA, C.-X.; WU, S. S.; ZENG, Z.-B. Linkage mapping of sex-specific differences. **Genetical research**, v. 79, n. 1, p. 85-96, 2002b.

YANG, G. P.; SAGHAI MAROOF, M. A.; XU, C. G.; ZHANG, Q.; BIYASHEV, R. M. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. **MGG Molecular & General Genetics**, v. 245, n. 2, p. 187-194, 1994.

YU, J.; HU, S.; WANG, J.; et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). **Science**, v. 296, p. 79-92, 2002.